ISPA | Instituto Superior de Psicologia Aplicada

Mestrado em Etologia

# Cooperation in rats playing an Iterated Prisoner's Dilemma game: influence of a game matrix formed with qualitatively distinct payoffs.



Patrício Manuel Vieira Simões

2007

ISPA | Instituto Superior de Psicologia Aplicada

**Mestrado em Etologia**

# Cooperation in rats playing an Iterated Prisoner's Dilemma game: influence of a game matrix formed with qualitatively distinct payoffs.



Dissertação orientada por:

Marta Moita, PhD (Instituto Gulbenkian Ciência)

Patrício Manuel Vieira Simões

**2007**

Este trabalho não inclui as observações feitas pelo júri.

**Index**

**Resumo**

O Dilema do Prisioneiro Reiterado (repetido) (DPR) é o paradigma central no estudo da cooperação de animais não-humanos. Este jogo formaliza os requerimentos descritos por Trivers (1971) para que a cooperação surja e se mantenha por reciprocidade. Neste jogo dois jogadores podem escolher numa jogada entre cooperar ou não cooperar: A cooperação mútua fornece a ambos os jogadores um reforço R (Reward), enquanto a não cooperação mútua fornece um reforço P (Punishment). Se um dos jogadores cooperar e o seu adversário não o fizer, o primeiro recebe um reforço S (Sucker) e o último recebe T (Temptation). Os reforços deste jogo terão que seguir as inequações T>R>P>S e 2R>T+S. O dilema deste jogo surge do facto de independentemente do que o adversário fizer, a escolha de um jogador que produz um maior reforço é não cooperar (T>R e P>S). No entanto ambos os jogadores receberiam um reforço maior se ambos cooperassem. Ser reiterado significa que existem um número não especificado de jogadas em que o resultado de uma determinada jogada poderá ser influenciado pelo resultado das anteriores. Axelrod e Hamilton (1981) mostraram que a cooperação poderá tornar-se sustentada num jogo DPR e que a estratégia Tit-For-Tat (TFT), em particular, é uma solução robusta para este problema. A estratégia TFT comanda um jogador a cooperar no encontro inicial e em encontros posteriores a copiar a decisão anterior do adversário.

Apesar do grande sucesso ao nível da investigação teórica, existem poucos dados empíricos que suportem a reciprocidade (e por consequência o DPR) como a explicação principal para a cooperação nos animais. De facto, e ao nível de estudos laboratoriais, animais sujeitos a uma matriz de jogo conforme ao DPR mostraram valores diminutos de cooperação. O insucesso em manter a cooperação através de um paradigma DPR experimentalmente controlado, levou os investigadores a questionar que mecanismos poderão prevenir a emergência da cooperação nestes moldes. Uma abordagem alternativa utilizada foi experimentalmente controlar um dos jogadores num jogo do DPR, sendo na maior parte dos casos usada a estratégia TFT. Estes estudos mostraram que os animais sob o paradigma DPR dão prioridade às consequências de curto prazo, enquanto desvalorizam o resultado de jogadas futuras. Para mais, as contingências de reforço, quer locais, quer passadas (como por exemplo, a magnitude de reforço entre os resultados T, R, P e S) podem modificar a probabilidade de cooperação do animal.

Partindo do princípio que jogar o DPR pode ser considerado uma tarefa de condicionamento operante, Stephens e Clements (1998) desenvolveram um modelo teórico que explora a relação entre os processos de aprendizagem e os equilíbrios teóricos do jogo utilizando matrizes de jogo com reforços positivos (recompensas), reforços negativos (castigos) ou ambos. A grande maioria dos estudos laboratoriais no DPR utiliza uma matriz de reforços positivos (na prática unidades de comida ou de dinheiro, se em humanos). No entanto este modelo apresenta um cenário interessante que deriva da aplicação de uma matriz de jogo em que os reforços S e P sejam: i) universalmente e sem ambiguidade considerados castigos e ii)

qualitativamente distintos dos reforços T e R. Usando este tipo de matriz heterogénea, o modelo de Stephens e Clements prevê que os jogadores exibam elevados níveis de cooperação (entre os 60% e os 100%, dependendo da taxa de aprendizagem). Segundo o nosso conhecimento, tal matriz nunca foi aplicada em estudos de cooperação em animais não-humanos.

Uma abordagem exclusivamente económica do DPR parece insuficiente para explicar a divergência entre predições teóricas e dados empíricos, visto que a estrutura clássica deste jogo não captura a sofisticação cognitiva que parece estar envolvida na cooperação por reciprocidade. Em concreto, os animais deverão ter capacidade de reconhecer o seu adversário como um indivíduo, compensar a diminuição do valor de reforço futuros e ter capacidade de memória suficiente para cumprir obrigações recíprocas de modo a que a cooperação por reciprocidade se mantenha. Para além dos obstáculos cognitivos, também as interacções sociais em si mesmas poderão limitar ou estimular a cooperação visto que a simples presença de um conspecífico poderá ter um valor de reforço não nulo. Deste modo, os reforços que os experimentadores tencionam dar numa experiência podem ser totalmente discordantes dos experienciados pelos animais.

Esta dissertação pretende examinar as escolhas de ratazanas quando sujeitas a um jogo do DPR utilizando uma matriz de jogo constituída por reforços positivos (T- 4 pepitas de chocolate e R- 1 pepita de chocolate) e reforços negativos (P- 1 beliscão na cauda e S- 3 beliscões na cauda). Os animais jogaram contra um conspecífico programado para responder segundo uma estratégia TFT ou uma estratégia aleatória. As tendências cooperativas e as estratégias globais dos animais foram analisadas e os possíveis constrangimentos cognitivos e sociais que possam explicar as observações foram discutidos.

Os resultados mostraram que quando as decisões são reciprocadas (oponente TFT), as ratazanas mostram níveis de cooperação sustentada de aproximadamente 60% por sessão, significativamente maior que os 12% de cooperação observada em ratazanas que jogaram contra uma estratégia aleatória. O sujeitos que jogaram contra um oponetne TFT parecem ter adoptado uma estratégia de jogo sub-óptima, mostrando níveis altos de cooperação sustentada e de não cooperação sustentada e alta probabilidade de "perdão". Ratazanas que jogaram contra uma estratégia aleatória pareceram mostrar uma estratégia de jogo a tender para o óptimo. Observou-se que a presença e posição de um conspecífico influenciou as decisões das ratazanas quando estas envolviam dois reforços positivos de diferente valor. Esta influência social foi nula quando a decisão envolvia dois reforços negativos. Este reforço externo proveniente da componente social dos animais poderá modificar o valor da matriz de jogo recebida pelos animais num contexto de DPR. Foi observada uma preferência por parte das ratazanas em entrar nos compartimentos imediatamente adjacentes ao conspecífico quando os reforços entregues eram positivos.Esta influência externa poderá modificar as têndencias cooperativas e as estratégias de jogo das ratazanas sujeitas a um jogo DPR. Estes animais mostraram-se sensíveis a contingências de reforço passadas e cooperaram significativamente

menos contra uma estratégia TFT (cerca de 26%) quando previamente submetidas a uma estratégia aleatória.

Este estudo demonstra que animais não-humanos podem apresentar altos níveis de cooperação sustentada num contexto DPR cuja matriz de jogo seja constituída por reforços positivos e negativos. No entanto, quer os níveis de cooperação observados, quer as estratégias adoptadas pelas ratazanas contra um oponente conspecífico são influenciados por efeitos sociais e de contingências de reforço passadas. Tal implica que uma perspectiva exclusivamente económica é insuficiente para explicar comportamentos cooperativos em animais. Não obstante, o DPR representa ainda uma ferramenta válida para o estudo da cooperação se se considerar os efeitos sociais e históricos próprios das interacções cooperativas entre os animais.

**Keywords:** Cooperação; Reciprocidade; Teoria de Jogos; Dilema do Prisioneiro; Tit-For-Tat.

**Abstract**

The Iterated Prisoner's Dilemma game (IPD) is the central paradigm in the study of non-kin cooperation through reciprocity in non-human animals. Despite great success of theoretical investigation, laboratory experiments using an IPD design revealed that animals showed low levels of cooperation. When playing against a Tit-For-Tat (TFT) strategy animals responded sub-optimally and devaluate the value of future outcomes. However, local and past reinforcement contingencies could modify the animal's cooperation probability. Stephens and Clements (1998) proposed that high levels of cooperation could arise if animals experienced an IPD in which the game matrix is formed by qualitatively distinct payoffs.

This study examined the cooperative tendencies and trial-by-trial strategies of rats playing against a conspecific programmed to play TFT or Random strategy in an IPD context with a game matrix constituted by positive (chocolate chips) and negative (tail pinches) reinforcements. The possible constrains that could explain the observations, and particularly the possible reinforcement value of social interactions, were discussed.

Results showed that when decisions are reciprocated, rats can sustain cooperation around 60% per session, significantly higher than 12% of cooperation observed in rats playing against Random. Rats that played against a TFT opponent seems to adopt a sub-optimal strategy across sessions, showing high indicators of sustained cooperation and defection and high levels of "forgiveness". Subjects that played against a Pseudorandom strategy tend to adopt an optimal strategy. Subjects showed a preference to stay in the adjacent room of the conspecific when the delivered reinforcements were positive. Such external influence could modify the cooperative tendencies and game strategy of rats playing an IPD game. Rats were also sensible to past contingencies and cooperate significantly less against TFT (around 26%) when previously submitted to random strategy.

This study demonstrates that animals could cooperate in an IPD with a game matrix composed with qualitatively distinct reinforcements. However, sustained cooperation and game strategies are affected by social effects and past contingencies. Such implies that an exclusively economical perspective seems insufficient to explain cooperative behaviours in animals. IPD could still represent a valid tool for study animal cooperation if taking in account possible social and historical contingencies intrinsic to cooperative interactions between animals.

**Keywords:** Cooperation; Reciprocity; Game Theory; Prisoner's Dilemma; Tit-For-Tat.

*A hydrogen bomb is an example of mankind's enormous capacity for friendly cooperation. Its construction requires an intricate network of human teams, all working with single-minded devotion toward a common goal. Let us pause and savor the glow of self-congratulation we deserve for belonging to such an intelligent and sociable species.*

Robert S. Bigelow, 1969, The Dawn Warriors

Human societies are based on large-scale cooperation among genetically unrelated individuals. Throughout life we help to generate social resources, from freeways to the Internet, via collective endeavours and divisions of labour. But such cooperation is always permeated by conflicts of interest, because resources create opportunities to compete and to cheat.

The ubiquity of human cooperation has drawn much attention from scientific, social and political thinkers throughout History. To illustrate this point, Dugatkin (1997, 2002) compiled several interesting historical episodes centred on the problem of human cooperation. Two of the best known are presented: In Leviathan, the 16$^{th}$ century British philosopher Thomas Hobbes argued that one critical function of government is to enforce cooperation among its constituents. Without such enforcement, man's nature would surely lead to a "war of all against all." Hobbes did not, however, believe that cooperation was unnatural to other species and in fact, devoted a number of pages of the Leviathan to why social insects are cooperative, and man is not (at least, not by nature). In contrast, Peter Kropotkin, one of Russia's foremost anarchists and a natural historian, saw cooperation all around him, and urged humans to renounce the "Hobbesian War":

"Don't compete! . . . That is the watchword, which comes to us from the bush, the forest, the river, the ocean. Therefore combine – practice mutual aid! That is what Nature teaches us; ... this is what man – the most primitive man –has been doing; and that is why man has reached the position upon which we now stand." (Kropotkin, 1902).

Current research in human cooperation is immense. Either using biological, psychological, economical or sociological approaches, researchers of these fields try to disentangle the features that make human cooperation unique. The classic mechanisms, centred in the genetic relations, repeated interactions or mutually beneficial contexts, are clearly insufficient to explain the phenomenon of human cooperation. In the other hand, the cognitive, linguistic and even physical human capacities are possibly in the biological base of human cooperation. These capacities allow the formulation of general norms of social conduct, the emergence of social institutions regulating this conduct, the psychological capacity to internalize

norms, and the basing of group membership on such non-kin characteristics as ethnicity and linguistic behaviour (Bowles and Gintis, 2003). The building of human norms and social institutions capable of rewarding cooperative agents and sanction non-cooperative agents are possibly based in key individual behavioural traits as prosocial emotions and strong reciprocity (Henrich and Henrich, 2006).

Strong reciprocity is a predisposition to cooperate with others and to punish those who violate the norms of cooperation, at personal cost, even when such costs will unlikely be repaid. A strong reciprocator responds to cooperative behaviour on the part of others by maintaining or increasing his level of cooperation, and responds to free-riding behaviour on the part of others by retaliating against the offenders even at a personal cost. The strong reciprocator is thus both a conditionally altruistic cooperator and a conditionally altruistic punisher whose actions benefit other group members at a personal cost (Ginnis et al., 2003).

Some powerful determinants of human cooperative behaviour are the cultural and genetic variability among human groups (Bowles and Gintis, 2003; Gintis, 2000), the structure of human social institutions (Fehr and Fischbacher, 2004), investment in reputation (Semmann, Krambeck and Milinski, 2004) and psychological (Levine, 1998; Kiyonari, Tanida and Yamagishi, 2000) or physiological (Rilling et al., 2002) capacities to internalize norms. Several studies tried to capture the subtleties of the interactions between selfish and reciprocal individuals and to understand the nature of the evolutionary forces that shaped human altruism. One important limitation to these studies lays in the fact that most of these experiments use students from developed countries as subjects, instead of subjects that are representative of whole countries or cultures. In consequence, there is a still a lack of knowledge of how behavioural measures of human altruism are related with demographic data and cultural and economic institutions. The role and importance of group conflicts, group reputation and the evolution of cooperation in larger human groups (larger than 2) remain open questions (Fehr and Fischbacher, 2004).

But cooperation is not an exclusively human feature, as illustrated by vampire bats that regurgitate blood to others despite the possibility of dying if three days elapse without consuming blood (Wilkinson, 1984); ground squirrels that give alarm calls even though they alert predators to their own presence (Sherman, 1977, 1985); or cleaner fish that enter in the mouths of their hosts to remove parasites even at risk of being eaten (Grutter, 1999). Such examples of animal cooperation have generated a substantial amount of theoretical and empirical work on the adaptive role of cooperative behaviours. To understand the circumstances that favour and sustain cooperation, especially when cheating can yield an immediate higher payoff, is not yet fully resolved and remains a much discussed question in biological and psychological sciences.

Within evolutionary biology, interest in animal cooperation can be traced back to Darwin (1859), who feared that the self-sacrificial behaviour of the eusocial insects (e.g, ants, bees, and wasps) was a threat to his entire theory of natural selection. In the first half of the twentieth

century, biologists collected a large amount of data documenting cooperation among animals but no theory was developed (Dugatkin, 1997). The theoretical framework to explain animal cooperation began with Hamilton's (1964) kin-selected cooperation and was followed by other three ground-breathing theories: reciprocity (Trivers, 1971), group selection (Wilson, 1975) and by-product mutualism (Brown, 1983, West-Eberhard, 1975).

**Kin-selected cooperation**

Perhaps the best-studied form of cooperation is kin selection. Since Darwin (1859) it is recognized that individuals are more prone to cooperate with relatives. Hamilton (1964) mathematically formalized this idea in what has become known as inclusive fitness or kin-selection model. At the core of inclusive fitness is that a cooperative or altruistic behaviour could be maintained toward their genetic relatives because it helps propagate their own genes. What seems altruistic from an individual's perspective actually serves self-interest from the gene's point of view.

One of the best known examples of kinship-based cooperation is Sherman's (1977, 1985) work on alarm calls in Belding's ground squirrels (*Spermophilus beldingi*). In this species, individuals often give alarm calls when a predator is sighted; such calls alert all those in the vicinity. Female, but not male squirrels give alarm calls when a predator is in the vicinity more often than expected by chance. Females are generally sedentary and breed near their natal sites, whereas males always emigrate from their birthplace and do not aggregate with siblings after emigration. Therefore, females' calls probably warn close kin (often offspring), whereas males' calls do not. Furthermore, invading (non-native) females give alarm calls less frequently than native females, and that females living relatives are more likely to call than are females without any living kin. These findings supports the hypothesis that these alarm calls serves to warn relatives rather than to solicit aid from others.

**By-product mutualism**

The simplest explanation for cooperative behaviour is that it provides direct benefits not only to the cooperator, but also to other individuals. Any individual that defects (i.e., does not cooperate) in mutualistic situations will, by definition, do worse than a cooperator; therefore, in the absence of a temptation to defect, cooperation provides the best option. That is, in by-product mutualism, the immediate net benefit of cooperating outweighs that of cheating (Brown, 1983, West Eberhard 1975). Since mutualism does not depend on the identity of your partner it can occur between any members of the same species and even across *taxa* (Dugatkin, 1997).

By-product mutualism may be the most common path to cooperation, but some behavioural ecologists argue that this mechanism should not be considered cooperation since in mutualistic scenarios cheaters receive a lower payoff than cooperators, and hence, there is no temptation to cheat (Brown 1983, Connor, 1995a).

Pied wagtail birds (*Motacilla alba*) defend riverside winter territories, foraging on insects and searching their territory systematically for new food items. Intruder wagtails often land on such territories and pose a threat to territory owners. Sometimes such intruders are tolerated, in which case they are termed "satellites"; at other times they are aggressively chased off a territory. If an owner allows an intruder to stay, it loses some foraging-related benefits, but it gains assistance in territory defense. On days of low food abundance, however, owners actively evict any intruders attempting to feed on their territory. Allowing the presence of satellites is beneficial to owners in days of food abundance, since this is when intruder value is highest and the associated cost is lowest. On the other side, intruder birds benefit from acting as satellites and engaging in cooperative territorial defense, since they get some food when owners allow entrance onto territories. Cooperation in both intruding and territory-holding wagtails is best understood as an example of by-product mutualism, since both engage in cooperative behaviours when they get a direct benefit (Davies and Houston 1981).

**Group-selected cooperation**

Trait-group selection contends that natural selection operates at two levels: within- and between-trait groups, where a trait-group is defined as all individuals that affect each other's fitness. Within-group selection is very similar to the "selfish gene" hypothesis (Dugatkin, 2002). That is, within-group selection acts against cooperators, since in trait group models, cooperators take on some cost that others do not. Defectors are always favoured by within group selection since they receive any benefits that accrue because of the actions of cooperators, but pay none of the costs. As opposed to within-group selection, between-group selection favours cooperation, if groups with more cooperators outproduce other groups. For example, when individuals in a group give alarm calls, they may pay a cost within groups (as they may be the most obvious target of a predator), but groups with many alarm callers may outproduce groups with fewer cooperators. The relative strength of within- and between-group forces determines whether cooperation evolves in any particular situation (Wilson, 1975, Wilson and Sober 1994).

An interesting putative case of trait-group selected cooperation has been documented during cooperative colony foundation in the seed harvester ant, *Messor pergandei*. Colonies are usually initiated by multiple unrelated queens. Adult colonies are very territorial, and brood raiding is seen among young starting colonies in the laboratory. Brood captured by nearby colonies are raised within the victorious nests, and colonies that lose their brood in such interactions die out. All queens produce workers, and there exists a positive relationship between the number of queen foundresses in a colony and the number of initial workers (which can act as brood raiders) produced by that colony. That is, between-queen cooperation increases group productivity, which translates into greater success at brood raiding. This example provides all of the elements needed for group selection to operate. Group selection requires the differential productivity of groups based on some trait. In the case of *M. pergandei*, the trait of interest is queen-queen cooperation, which, although selected against within groups

(a queen that did not cooperate might reap some benefits from her cooperative cofoundresses without paying any costs), may be selected for because groups with many cooperators survive brood raiding (Rissing and Pollock 1991).

**Reciprocity**

In his now classic paper, "The evolution of reciprocal altruism", Trivers (1971) argued that genes for cooperative and altruistic acts might be selected if individuals differentially display such behaviours to others that have already been cooperative and altruistic toward them. Essentially, a short-term cost of cooperation could be beneficial if, sometime in the future, the social partner's reciprocate cooperation. The requirements for stable evolution of reciprocity are: i) the reciprocated benefit must outweigh the immediate cost, ii) individuals must interact repeatedly, and iii) individuals must recognize each other. However, such system could be subject to cheating when the greatest payoff attainable goes to the recipient of a cooperative action who fails to reciprocate in turn. How then might reciprocity evolve in such a model?

One way to examine this question is through the use of a mathematical game called the Prisoner's Dilemma (PD). The general aspects of the PD are: i) cooperation maximizes the total payoff to everyone involved in the interaction (mutual cooperation provides more benefits than mutual defection); ii) any individual will receive a higher personal payoff by defecting, so a sizable temptation to cheat exists (Trivers, 1971). The PD is a particular game of the Game Theory, a mathematical and economic tool that has also proven to be a powerful predictive tool for behavioural and evolutionary ecologists in several biological problems (Dugatkin and Reeve, 1998).

A well-cited example of reciprocity is Wilkinson's (1984) work on blood sharing among vampire bats (*Desmodus rotundus*). These bats typically live in groups composed largely of females, with a low average coefficient of relatedness (between 0.02 and 0.11). Remarkably, females in a nest of vampire bats regurgitate blood meals to others that have failed to obtain food in the recent past. This sort of food sharing can literally be a matter of life or death, as individuals may starve if they do not receive a blood meal every 60 h. While kinship plays a role in reciprocity and food sharing in vampires, Wilkinson suggests two strong lines of evidence that reciprocity, above and beyond any effect of relatedness, is important in vampire blood sharing: i) the blood meal obtained is critical, while the cost of giving up some blood may not be that great, thus satisfying one of the conditions stipulated by Triver's model of reciprocal altruism; ii) vampires are able to recognize one another and are more likely to give blood to those that have donated in the past.

**Defining cooperation**

But what is cooperation? Most of us will have an intuitive idea of what is meant by "cooperation", but the definition is not straightforward as one may assume. Cooperation research attracted interest from scholars of many different disciplines, some of which concentrate on human behaviour, while others focus on interactions among nonhuman organisms. An exchange of ideas, which is no doubt in the interest of all disciplines, is hindered by a confusion of language like a Babel tower with different traditions, different jargon and different goals: within biological research the most commonly used terms (cooperation, mutualism, reciprocity, reciprocal altruism and symbiosis) are sometimes used as synonyms, sometimes as categories in a hierarchical system and sometimes to denote mutually exclusive categories (Nöe, 2006).

With respect to the definition of cooperation, several terminologies were proposed (Connor, 1995b; Nöe, 2006; Sachs et al., 2004; Stephens and Anderson, 1997), in which two confronting perspectives arise: i) a phenomenological definition of cooperation, focusing on the elaborate behavioural coordination of individuals that seem especially sensitive to the actions of others, as is it implicit in "cooperative hunting" or "cooperative breeding" terms (Roberts, 1997), or ii) an operational definition that de-emphasize the behavioural properties of an interaction, focusing on the economic consequences of the behaviours (Stephens and Anderson, 1997) and commonly generalized as a "joint action for mutual benefit" (Dugatkin et al. 1992). The last definition has increasingly been adopted since the integration of game theory tools into the research of animal cooperation because it can be objectively applied and looks beyond casual and elaborate cooperative interactions. In that sense, cooperation could be seen as "an outcome, not a mechanism": the outcome of cooperation may be achieved by many different mechanisms, such as kin selection, reciprocity or mutualism (Stephens and Anderson, 1997).

In the latest terminology proposed for cooperation research terms, Nöe (2006) advocated that cooperation as a "joint action for mutual benefit" is a definition build under highly artificial experiments and procedures used to study behavioural strategies specific to cooperation and test game-theoretical models. In these cases the impression of cooperation is created because animals are tested after having been trained to interact with an apparatus or because there was an uncontrolled flow of information. He labelled these phenomenons, respectively as "instrumental cooperation" and "communicative cooperation".

**Game Theory**

Game theory studies decisions that are made in an environment where various players interact and the costs and benefits of a decision depend upon the choices of other individuals. First developed as a tool for understanding economic behaviour and then by the RAND Corporation to define nuclear strategies, game theory is now a widespread research tool in diverse fields such as biology, psychology, sociology, philosophy, computer, military and political sciences. Within biological research, Game Theory has being used to address a broad

range of problems such as social foraging, animal aggression and conflict, communication, sexual selection, parent-offspring conflict, habitat selection, predator-prey dynamics, learning and quantitative genetics (Dugatkin and Reeve, 1998).

A game is a well defined mathematical object that essentially consists in a set of players, a set of moves (or strategies) available to those players, and a specification of payoffs for each combination of strategies often represented in a matrix. The different relations possible between number of players, strategies and payoffs define a particular game object. Several other attributes could be used to increase complexity of games: These could be symmetric, if the payoffs for playing a particular strategy depend only on the other strategies employed and not on who is playing them, that is, if the identities of the players can be changed without changing the payoff to the strategies, or asymmetric games where there are not identical strategy sets for both players (Osborne and Rubinstein, 1994).

A game could be a zero-sum game if the total benefit to all players in the game, for every combination of strategies always adds to zero (or more informally put, a player benefits only at the expense of others) or non-zero-sum games when some outcomes have net results greater or less than zero (informally, in non-zero-sum games, a gain by one player does not necessarily correspond with a loss by another) (Osborne and Rubinstein, 1994).

Games can also be simultaneous if players move simultaneously, or if they do not move simultaneously the later players are unaware of the earlier players' actions (making them effectively simultaneous). Sequential games (or dynamic games) are games where later players have some knowledge about earlier actions. This need not be perfect knowledge about every action of earlier players; it might be very little information. For instance, a player may know that an earlier player did not perform one particular action, while she does not know which of the other available actions the first player actually performed. An important subset of sequential games consists of games of perfect information. A game is one of perfect information if all players know the moves previously made by all other players. Thus, only sequential games can be games of perfect information. Most games studied in game theory are perfect information games (Osborne and Rubinstein, 1994)

**Prisoner's Dilemma game**

The most famous game in Game Theory is undoubtedly the Prisoner's Dilemma (PD) problem (Fig.1). In this game, two players can choose to either cooperate or defect: Mutual cooperation provides both players with a payoff labelled R (Reward), while mutual defection produces a payoff of P (Punishment) for both players. Should one player cooperate, but its partner defect, the former receives payoff S (Sucker), while the latter receives T (Temptation). The payoffs of this game are set up as follows: T>R>P>S (another inequality, 2R>T+S, is often tacked on). The dilemma in this game centers on the fact that, regardless of what a player's opponent does, it receives a higher payoff when defecting (as T>R and P>S). This is true for both players in a game and hence both individuals should defect and end up with a payoff of P.

Yet both players would had received a higher payoff if they had both cooperated (as R>P). On a single play of PD, then, rationality as well as natural selection leads players toward a suboptimal solution (Dugatkin and Reeve, 1998).

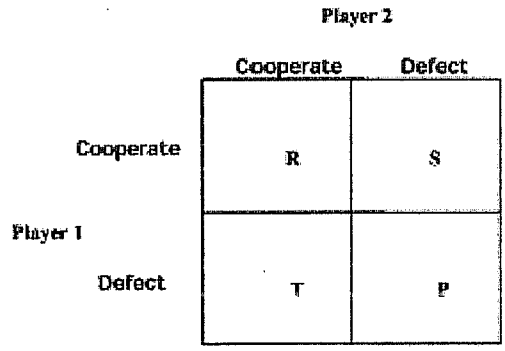


**Figure 1.** The prisoner's dilemma game matrix. Each player can each choose to cooperate or defect (not cooperate). The game is defined by the inequality T > R > P > S. Payoffs are shown for Player 1.

**Iterated Prisoner's Dilemma**

Is there any escape from this dilemma? That is, can cooperation ever evolve in such a scenario? Axelrod and Hamilton (1981), using both analytical techniques and computer simulations, examined the success of an array of behavioural strategies in the Iterated Prisoner's Dilemma (IPD) game, where an iterated game is one that is played over and over between the same individuals. They found that if the probability of meeting a given partner in the future was above some critical level, then conditionally cooperative strategies could arise and dominate the simple strategy of "always defect". Based on these simulations Axelrod (1984) organized an IPD tournament and invited academic colleagues all over the world to devise computer strategies to compete. The programs that were entered varied widely in algorithmic complexity; initial hostility; capacity for forgiveness; and so forth. The winner of the tournament was Anatol Rapoport's Tit For Tat (TFT) strategy. TFT is a simple strategy that instructs a player to cooperate on the initial encounter with a partner and, in subsequent encounters, to copy the partner's last move. Axelrod (1984) hypothesized that TFT's was a robust solution to the IPD and that could be attributable to its three defining characteristics: "niceness" (someone employing TFT is never the first to defect), swift "retaliation" (individuals using TFT immediately respond by defecting, if their partner cheats), and "forgiving" (a partner using TFT remembers only one move back in time and therefore forgives prior defection on the part of a currently cooperating partner - that is, a partner using TFT does not hold grudges).

Following intense theoretical studies in diverse IPD scenarios, Nowak and Sigmund (1993) presented a strategy that in some conditions could outperform TFT strategy: PAVLOV. This strategy could be informally described as a win-stay, lose-shift (WSLS) strategy. This family of strategies (PAVLOV is one of many WSLS possibilities) is defined by an initial action

and a certain "aspiration-level", i.e., a minimum expectation concerning the payoff in each round. If this expectation is satisfied the players stick to the current behaviour, if not, they switch to the alternative action. The aspiration level of the PAVLOV strategy for example lies between the cooperative payoff and the payoff for mutual defection. If the aspiration level is below all possible payoffs the WSLS rule will always repeat the initial action (Posch, 1999). Although successful in simulation scenarios, WSLS strategies are extremely difficult to implement in empirical studies in non-human animals due to the difficult to control or measure the expectation level of the subjects along the experiments, or in simple words, to know if an animal experienced a victory or a defeat in a given payoff (Stephens and Clements, 1998).

**Classical ethological research in animal reciprocity**

Following of Axelrod and Hamilton's (1981) IPD model, evolutionary theorists produced an overwhelming amount of articles adding an increased complexity and sophistication to the reciprocity research: modifications to the game include variations in number of players, population structure, environment dynamics, strategy stability, amount of memory, norms, mobility of players, mistakes or payoff structure (For a review on these studies, see e.g. Dugatkin and Reeve, 1998; Doebeli and Hauert, 2005). On the other hand, empirical studies centred in the IPD model did not follow the theoretical research: in fact, only a handful of species have demonstrated reciprocity, and even within these species reciprocity occurs infrequently or alternative explanations to the behaviour could be vented (Dugatkin, 2002). Many authors have reported reciprocity in numerous contexts including food sharing in primates (de Waal and Berger 2000, Hauser et al. 2003) and bats (Wilkinson, 1984), grooming in blue monkeys (Rowell et al. 1991) and impalas (Hart and Hart 1992), predator inspection in fishes (Dugatkin 1988, Milinski 1987), or coalitions associated with mating opportunities in baboons (Packer, 1977). Unfortunately, most examples of reciprocity suffer from one of two problems: i) they have not been replicated, and ii) alternative explanations, such as kin selection and mutualism, can account for the observed reciprocal pattern (Stevens, Cushman and Hauser 2005).

One of the most cited field studies of a behaviour that appears to involve some sort of reciprocity is the allogrooming behaviour in impala (*Aepyceros melampus*) (Hart and Hart 1992). Bouts of reciprocal allogrooming are initiated when one impala begins grooming another. Although in some species grooming is solicited by the individual to be groomed, in impala it is the groomer who initiates this activity. Each bout of allogrooming consists of the initiator grooming the recipient 6-12 times and the recipient responding in kind. A typical exchange involves three to four such bouts. Grooming reduces tick infestation, which left untreated can have detrimental effects on impala health. The extent of reciprocity within grooming pairs of impala is impressive. Whether pairs are male-male, female-female, female-male, sub-adult male-subadult male, or fawn-fawn, each partner receives almost the same number of allogrooming bouts that it hands out. The fact that reciprocal allogrooming is seen in fawns as

young as three days old, and that young deprived of allogrooming exchanges with adults still display this behaviour, suggests that there has been strong selection pressure for reciprocal allogrooming.

On the other hand, reciprocity in predator inspection on guppies (*Poecilia reticulata*) (Dugatkin, 1988) gave rise to great controversy about the true mechanism involved in this behaviour. In many fish species, one or a few individuals breaks away from a school and approaches a predator to assess the potential danger. The costs and benefits of predator inspection probably meet the criteria for a prisoner's dilemma. In other words, the best option for a fish is to stay back and watch its partner inspect (T > R), but if both fishes fail to inspect (and receive P), then they are likely worse off than had they both approached the predator (and received R), and thus T > R > P > S. Results of Dugatkin's experiment (1988) and posterior researchers (Dugatkin 1991a, b, Milinski 1996) suggest the notion that guppies reciprocate predator inspections in a TFT-like strategy. That is, inspectors appear to be nice (each starts inspecting at approximately the same time), retaliatory (inspectors cease inspecting if their partner stops), and forgiving (if inspector A's partner has cheated it in the past but resumes inspection, then A resumes inspection as well). In addition to this direct support for TFT, evidence exists that inspectors remember the identity and behaviour of their co-inspectors and prefer to associate with cooperators rather than cheaters (Dugatkin and Alfieri 1991). These experiments elicited a flood of criticisms: in particular, that the fitness payoffs of cooperation and defection were unclear (Connor 1996, Lazarus and Metcalfe 1990), and rather than reciprocating, the target fish may simply have preferred to stay in groups to reduce predation risk (Masters and Waite 1990, Stephens et al. 1997). This controversy is still today unresolved.

**The psychological approach to non-human cooperation**

The classic ethological approach on the study of non-human animal reciprocity illustrated by the studies cited above share a common criticism: there is no experimental control and is difficult or impossible to quantify the real value of the payoffs implicated in the game matrix. In consequence, a researcher facing a given cooperative behaviour can only infer that it's a PD situation and thus that a reciprocity mechanism driven such behaviour (Connor 1996; Nöe, 2006), in what its being known as the first wave of evolutionary economics research in biology. In alternative, a cognitive ethological and experimental psychology approach as been suggested by some researchers to entangle the reciprocity mechanisms in non-human animals with laboratory experiment designs centred in the experimental control of reinforcements, which can be interpreted as an effective control over the payoff matrix. In this perspective, the focus is centred in the economical outputs of the animal's actions under a controlled game context. This was facilitated by the recent increase in the interdisciplinary discourse and abstract tool share between economist and biologists (Hammerstein and Hagen, 2005).

The study of the cooperative mechanisms in controlled payoff contexts in non-human animals took the first steps within a strictly "behaviourist" framework since playing IPD game

could be considered an operant conditioning task (Stephens and Clements, 1998). The most basic idea in operant conditioning is the Thorndike's (1911) "law of effect" in which responses associated with reward would increase in frequency, and responses associated with punishment would decrease in frequency. Albeit the criticism of circularity of this concept, the law of effect provided a useful organizing framework for operant psychology research. In particular, operant psychologists seek refinements of the law of effect in terms of the economical and psychophysical properties of stimuli and taking in account the contingencies ("predictive" links between action and consequences) necessary for conditioning to occur (Stephens and Clements, 1998).

Within this psychological and economical framework several researchers tested two players iterated games following the PD matrix in rats (Flood et al. 1983; Gardner et al., 1984), pigeons (Green et al., 1995), starlings (Reboreda and Kacelnik, 1993) and blue jays (Clements and Stephens, 1995; Stevens and Stephens, 2004). In all these studies the animals have shown low levels of cooperation, but accessory and complementary information about the psychological mechanisms of cooperation were unveiled.

Flood et al. (1983) examined the choices made by male rats under the differential reinforcement learning theory that hypothesized that in IPD, fixation was expected either in stable cooperation or stable defection. The authors used 2 adjacent operant chambers with coarse wire mesh between in which 2 rats could use a "cooperative" lever or a "defective" lever to access a food dispenser. The payoffs were defined as negative payoffs (different time delays until same food deliverance) to attempt to overcome satiation problems using food as positive reinforcements. The average response of the rats was around 70% of non-cooperative decisions in the final run. Albeit not concluding the existence of contingent strategies to rats, the authors found that there is no initial bias towards a final preferred strategy, which suggest a slow and incremental learning process.

Gardner et al. (1984) also used rats with a game theory methodology to investigate cooperative tendencies of pair of subjects playing IPD with and without visual information about the game partner. Two T-mazes displayed "face-to-face" were used to test female rats under a positive reinforcement matrix (food pellets). Cooperation or defection responses were assigned to the extreme of a given arm of the maze. The subject played simultaneously in a first phase with a Plexiglas partition, which was replaced by an opaque partition in a second phase. Results showed that with or without an opaque partition the most common behaviour was mutually defection response, which was accentuated when the visual communication was eliminated (60% and 74% of average mutual defection for transparent and opaque partition, respectively). Although a definitive statement about the effect of a barrier in the diminish of the cooperative responses cannot be made, since no control group was used, a measurable trend was found between the choices made by rats during the two phases that suggests that information about the partner's choices alter the contingencies under which the subject decide, or quoting the eloquent conclusion of the authors, the rat is essentially an uncooperative organism.

The failure of these experiments in eliciting sustained cooperation through an IPD paradigm lead researchers to use several protocol variations to investigate which constrains could explain the divergences between the empirical results and the theoretical assumptions. This alternative approach was mainly made using blue jays (*Cyanocitta cristata*) and pigeons as models. The common twist in this approach is the experimentally control the behaviour of one of the players in the IPD game, and in particularly, force such "stooge" player to act under a TFT strategy. Playing a prisoner's dilemma game against an opponent playing tit-for-tat is analogous to a test of self-control. In this situation, the alternatives on a given trial are a smaller versus a larger reinforcer, both immediate: choice of the smaller reinforcer leads to larger overall reward; choice of the larger reinforcer leads to smaller overall reward. In other words, the immediate consequence of defection is an increase in reinforcement rate, but its delayed consequence is a decrease in what is termed as "complex ambivalence" (Rachlin, 2000). Baker and Rachilin (2002) tested pigeons in an operant chamber programmed to respond as a TFT opponent in an IPD context. The payoff matrix, though being formed by positive values (time assessable to food), its relative magnitude was increased, that is, the distance between T and S was augmented (T=6>R=5>P=2>S=1) in relation with matrixes in previous studies. Results showed approximately 60% of cooperation choices in treatments with null inter trial intervals (ITI) and approximately 45% in treatments with 18 seconds ITI. The interaction between the experimental control and short duration of the ITI and the matrix payoff relations were suggested as the possible causes to the relatively high proportion of cooperative choices of pigeons.

Hall (2003), on the other hand using a similar methodology, but via a "classical" positive reinforcement matrix (T=5>R=3>P=1>S=0), tested pigeons playing against a TFT and a Random computer software. Prior to the tests, the birds were submitted to an altered matrix in which the equilibrium strategy was mutual cooperation (mutualistic matrix: R>T>P=S). Results indicated that choice allocation was optimal when the birds played against Random, but was sub-optimal when the birds played against TFT (that is, rapid diminish of the cooperation levels in both conditions). In order to determine why the pigeons responded suboptimally against TFT, a trial-by-trial analysis of the data revealed that once a pigeon had received the 'Sucker's' payoff (S), it was more likely to defect and receive the 'Temptation' payoff (T) than to cooperate and receive the 'Reward' (R) payoff. Local reinforcement contingencies appear to determine suboptimal responding against TFT in the IPD.

Stevens and Stephens (2004) tested blue jays in a series iterated games using a special feeding apparatus with visual communication to investigate the stability of cooperative choice against a conspecific "stooge" trained to follow different strategies. The experiment tested four game matrices, called "cooperate only" (R=S>T=P), "defect only" (T>R=P>S), "prisoner's dilemma" (T>R>P>S), and "opponent control" (T=R>P=S) treatments. This study found little cooperation in the "defect only" and IPD treatments (approximately 10% of mutual cooperation in each treatment) and occurred significantly more often in the "opponent control" treatment (around 60% of mutual cooperation). These findings suggest that the jays attend to short-term consequences; they do not cooperate in the absence of an immediate benefit (defect

only), even if a long-term benefit may exist (prisoner's dilemma). The opponent control treatment suggests that cooperation can occur when an individual's benefits depend completely on the actions of others; therefore, generosity is cheap.

Using a similar experimental design, Stephens et al. (2002, 2006) tested jays to play against a TFT opponent. The subjects were separated in several treatments to examine the role of payoff accumulation and temporal clumping (proxies to control animal impulsiveness towards immediate payoffs and the devaluation that the animals make towards delayed rewards). Results show high stable levels of cooperation in treatments where the animal impulsiveness were minimized (approximately 70% of final proportion of cooperation), significantly higher than in treatments where animals where free to receive immediate payoffs (around 50% proportion of cooperation). These recent experiments suggest that stable reciprocal cooperation requires that animals should receive the payoffs of a given behaviour inside a given temporal window to avoid value devaluation.

These recent studies have raised a series of questions focusing constrains that could justify the lack of reciprocal cooperation in laboratory investigations in non-human animals. The role of a straightforward IPD paradigm therefore seems insufficient to explain the divergence between theory and experiments.

**Psychological constraints on reciprocity**

A strictly adaptive perspective has limited power to explain the frequency of mutualistic and kin-biased cooperation mechanisms, and the rarity of reciprocity. A proximate perspective that keeps its eye on the ultimate problems can, however, reveal how psychological constraints could limit or facilitate particular forms of cooperation. This proximate approach emphasizes critical aspects of reciprocity that differ markedly from other mechanisms of cooperation (Stevens, Cushman and Hauser 2005).

In by-product mutualism cooperation, individuals should always cooperate since there no temptation to cheat exists. As a result, mutualism requires no special cognitive abilities above and beyond the challenges inherent in the cooperative behaviour itself (Stevens, Cushman and Hauser 2005). Kin-biased cooperation, on the other hand, does require additional cognitive capacities. At a minimum, it requires the capacity to direct cooperative actions to related individuals (Dugatkin and Alfieri 2001). Mechanisms of kin recognition include recognition alleles (individuals can compare a particular phenotypic cue to an innately specified template), phenotype matching (individuals compare a conspecific's phenotypic cues to a learned template) and familiarity cues (a simple set of spatial and rules and familiarity cues to discriminate kin, such as "be nice to individuals near your home" or "help those that you grew up with"). Of these three, the most sophisticated cognitive mechanism is phenotypic matching that requires specialized perceptual and computational systems that detect cues at an early stage to form a template, then test cues against the template to discriminate kin (Hauber and Sherman 2001).

In reciprocity, the fitness benefits associated with cooperation depend on the partner's behaviour: cooperation should only occur when the partner responds by reciprocating or punishing. When this contingent response occurs in the future, the temporal delay introduces cognitive challenges that may constrain the emergence and stability of cooperation. Stevens, Cushman and Hauser (2005) argued that the requirements formulated by Trivers' (1971) for the evolutionary stability of reciprocity do not capture the cognitive sophistication required for overcame such challenges and proposed 3 cognitive constraints on reciprocity that could explain the scarcity of cooperation by reciprocity in nature: i) individual recognition, ii) temporal discounting and iii) memory.

Individual recognition is needed as a cognitive ability to avoid cheaters and stabilize cooperation. However, since numerous species across the animal kingdom possess the ability to recognize individuals, the necessity for such mechanism cannot explain the paucity of cooperative behaviour across most nonhuman *taxa* (Stevens, Cushman and Hauser 2005).

Temporal discounting is the devaluing of future rewards, which often results in a preference for smaller, immediate rewards over larger, delayed rewards. The discounting problem could be considered analogous to the PD problem: Individuals must choose between the immediate reward of defecting and the long-term reward of cooperating (Stevens, Cushman and Hauser 2005). Estimated hyperbolic functions for temporal discounting using a adjusting-amount procedure in pigeons (Mazur 1987) and rats (Richards et al. 1997) and individual data points for tamarins and marmosets showed very high levels of impulsivity, that is, the value of a delayed reward decreases with the time to receiving the reward (Stevens, Cushman and Hauser 2005) (Fig.2). In contrast, humans show parallel reductions in value in months rather than seconds, even taking in account important differences between the human and nonhuman studies including the reward currency (money versus food) and experimental techniques (hypothetical situations versus operant training) (Rachlin et al. 1991). The temporal devaluation of rewards is correlated with lower level of cooperation both in humans (Harris and Madden 2002) and in blue jays (Stephens et al. 2002, 2006), suggesting that immediate benefits of defections may outweigh the future reciprocated benefits.

Limitations in memory decay, interference, and capacity can also constrain the frequency of reciprocity. Since memories decay over time, time intervals between cooperative acts may make reciprocity more difficult. Even with short time delays between cooperative interactions and few distractions, every potential new partner increases the computational load of tracking debts owed, favours given, and costs imposed. Keeping score of reciprocal obligations with multiple individuals may place a computationally intensive burden on memory systems (Stevens, Cushman and Hauser 2005). Although few studies examine learning and memory constraints in animal cooperation, Milinski and Wedekind (1998) study suggest that these memory constraints can pose challenges for maintaining stable cooperative relationships even in humans.

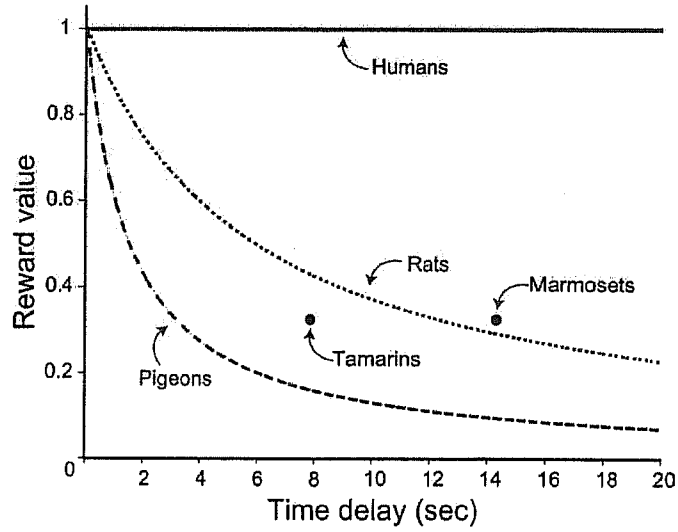


**Figure 2.** Species comparison of temporal discounting. The value of a reward decreases by 50% in the first 2–6 s in the nonhuman animals (in Stevens, Cushman and Hauser, 2005).

**External payoffs**

Another dimension that could constrain or enhance cooperation is the social interactions between animals. Animals can hardly be expected to abandon their natural behavioural repertoire as superfluous baggage once they enter the experimental apparatus since being with a conspecific can be a reward in itself. Such "external payoffs", or "intrinsic reinforcement" in the jargon of psychology, can cause considerable discrepancy between payoffs intended by the experimenters and payoffs as perceived by the subjects. The problem is that it is hard to quantify such payoffs and often not even know whether if such payoffs are positive or negative.

Most species used in cooperation experiments use "safety-in-numbers" as a defence against predators. Such species remain inclined to seek the vicinity of conspecifics even in a "safe" laboratorial environment and in the absence of predators. Seeing a conspecific in an otherwise stressful environment may therefore be a reward in itself (Nöe, 2006). Gardner et al. (1984) reported that rats cooperated better if they could see each other, a result that was confirmed for capuchin monkeys (Mendres and de Waal, 2000) and cotton-top tamarins (Hauser et al., 2003) in coordination tasks. Coordination is an important proximate mechanism needed to accomplish cooperation. By putting emphasis on this aspect of behaviour, theses studies resembles studies of rodents that have to coordinate movements in order to obtain a reward ("shuttling experiments ", rather than primate cooperation experiments. Schuster and Perelberg (2004) gave rats the choice to work together or to work alone for the same quantity of food and thus explicitly showed that rats prefer company. Subjects may also experience the presence of a conspecific as a negative reinforcer, because they see it as a competitor for food, territories or mates.

## 1. Introduction

Since Axelrod and Hamilton (1981) and Axelrod (1984) investigations in the theoretical evolutionary equilibrium of the Iterated (repeated) Prisoner's Dilemma (IPD), this game become the central paradigm in the study of non-kin cooperation in non-human animals. The IPD game formalizes the requirements described by Trivers (1971) for cooperation to arise and be maintained through reciprocity: i) the reciprocated benefit outweighs the immediate cost to cooperate; ii) individuals must interact repeatedly, and iii) individuals should recognize each other. Those studies motivated the economic examination of reciprocity by supporting the concept of Tit For Tat (TFT) - a simply strategy that instructs a player to cooperate on the initial encounter with a partner and, in subsequent encounters, to copy the partner's last move - as a robust solution to the IPD problem.

Despite the great success of the following theoretical investigation, adding an increased complexity and sophistication to the reciprocity research (for review, Dugatkin and Reeve, 1998; Doebeli and Hauert, 2005) there is little empirical evidence supporting reciprocity as a primary explanation for cooperation. In fact, only in a handful of species has been demonstrated reciprocity, and even within these species reciprocity occurs infrequently or alternative explanations to the behaviour could be vented (Dugatkin, 2002). Several laboratory experiments using non-human models focusing the economical outputs of the animal's actions under a controlled payoff matrix, tested two-players iterated games following the PD matrix: in rats (Flood et al. 1983; Gardner et al., 1984), pigeons (Green et al., 1995), starlings (Reboreda and Kacelnik, 1993) and blue jays (Clements and Stephens, 1995; Stevens and Stephens, 2004). In all these studies the animals showed low levels of cooperation.

The failure of eliciting sustained cooperation through an experimentally controlled IPD paradigm raised a series of questions that could justify the lack of reciprocal cooperation in laboratory investigations in non-human animals. Researchers used several protocol variations to investigate which constraints could explain these observations. This novel approach was mainly used in blue jays (Stevens and Stephens, 2004; Stephens et al., 2002, 2006) and pigeons (Baker and Rachilin, 2002; Hall, 2003) as models. The common twist in this approach is that the "opponent" in the game was actually a confederate (a conspecific or a computer program) of the experimenter, assigned to play in a particular strategy throughout the game. This intra-subject version of the game, allows an experimenter to pit the individual concerned against a known adversary, an appealing independent variable for investigators to manipulate. The most commonly studied adversary to date has been TFT strategy (Hall, 2003). These investigations revealed that animals attend to short-term consequences, devaluating the value of future outcomes (Stevens and Stephens, 2004; Baker and Rachilin, 2002) and that local (Hall, 2003) and past (Clements and Stephens, 1995) reinforcement contingencies could modify the animal's cooperation probability.

Playing an IPD game could be considered an operant conditioning task. Stephens and Clements (1998) developed a model that explored the relationship between operant

conditioning mechanisms and game theoretical equilibria. The basic idea of this model resides in the analysis of agents learning under an IPD matrix in which the payoffs could be positive reinforcements (rewards), negative reinforcements (punishments), or both. Laboratory experiments using IPD in non-human animals mainly used positive reinforcements (essentially food pellets) (e.g. Gardner et al., 1984; Clements and Stephens, 1995), although a few studies interpreted differential time until the access of food as negative reinforcements (e.g. Flood et al. 1983). Within this matrix design, Stephens and Clements model (1998) predicts that cooperation only becomes stable when agents present very high levels of learning rates and that in smaller learning rates mutual defection equilibrium is more "attractive" than the mutual cooperation equilibrium.

An interesting scenario also studied in this model is when Sucker and Punishment outcomes are: i) unambiguously and universally punishing and ii) qualitatively distinct from the Temptation and Reward consequences. In this kind of heterogeneous IPD matrix agents are predicted to show higher levels of cooperation (ranging 60% to 100% of sustained cooperation depending in the learning rate value) (Stephens and Clements, 1998). Early empirical experiments (Rapoport and Chammah, 1965) using such matrix design in humans showed high levels of cooperation and suggested that we treat losses and gains as qualitatively distinct entities, one of the cornerstones of the Prospect Theory of human decision making (Kahneman and Tversky, 1979). However, to our knowledge, no similar study was conducted in non-human animals.

A straightforward IPD paradigm seems insufficient to explain the divergence between theoretical predictions and empirical data, introducing the problem of the cognitive challenges that may constrain the emergence and stability of cooperation through reciprocity in non-human animals. Stevens, Cushman and Hauser (2005) argued that the requirements formulated by Trivers (1971) for the evolutionary stability of reciprocity do not capture the cognitive sophistication required for overcome such challenges and proposed 3 cognitive constraints on reciprocity that could explain the scarcity of cooperation by reciprocity in nature: i) individual recognition, ii) temporal discounting and iii) memory.

Another element that could constrain or enhance cooperation is the social interactions between animals. A simple analysis of the economical decisions made by the animals under a game context as predicted by the classical IPD paradigm fail to account for the possible reinforcement value of being with a conspecific. Such "external payoffs" can cause considerable discrepancy between payoffs intended by the experimenters and payoffs as perceived by the subjects. The problem is that it is hard to quantify such payoffs and often determine whether such payoffs are positive or negative (Nöe, 2006).

The present study utilizes a game-theory methodology to examine the choices made by male rats playing the Iterated Prisoner's Dilemma with a game matrix constituted by positive (T and R outcomes) and negative (P and S outcomes) reinforcements. The cooperative tendencies and trial-by-trial strategies of subjects presented to a conspecific programmed to play a TFT strategy were analysed and compared with a Random strategy. The possible constrains that

could explain the observed results were discussed. This study is part of a larger effort to understand the behavioural constraints on cooperation and to develop more plausible alternative models towards an improved IPD paradigm.

## 2. Materials and Methods

### 2.1. Subjects

The experiments were performed using experimentally naive male Sprague Dawley® rats (*Rattus norvegicus*). The animals were housed in pairs with free access to water and food. There were no restrains in their free-feeding weights. All animals were habituated to human manipulation during approximately a week. Since rats are neophobic to food (Galef and Whiskin, 2003) during this period the rats were also habituated to chocolate rice chips that corresponded to the positive reinforcements in the PD matrix.

Although Norway rats are not renowned for spontaneously cooperating under free-ranging condition, they are nevertheless a highly social and adaptable species that, in the wild, live in colonies characterized by dominance and a repertoire of social signals. Rats also exhibit social learning whereby feeding habits can transfer between individuals in a variety of ways. It may be that many species can readily learn to coordinate behaviours if there is sufficient sensitivity to the presence and behaviours of others (Galef and Whiskin, 2003).

### 2.2. Apparatus

Two T-mazes adjoined "side-by-side" and separated by a wire mesh wall as shown in Fig. 3 were used. The wire mesh partition increased "communication" between the rats, making the situation slightly more natural while knowingly sacrificing independent choice. The start boxes (20x6x7 cm) were made in plexiglas and communicated with the "arms" of the T-mazes trough a 7x7 cm orifice. Two rooms painted in black (20x20x20 cm each) and divided by a sliding wire mesh door corresponded to the "arms" of the T. The upper panel of each room was made of plexiglas to permit image recording. A recording camera and a 25W light were placed 50cm above the apparatus.

### 2.3. Game matrix

We assumed cooperation under the operational definition adopted by Stephens and Anderson (1997). In describing payoff matrices, we use the conventional Prisoner's Dilemma notation, that is, a focal player obtains R (the reward for cooperation) if both players cooperate, obtains S (the sucker's payoff) if the focal player cooperates and his opponent defects, obtains T (the temptation to cheat) if the focal player defects and his opponent cooperates, and obtains P (the punishment payoff) if both players defect. We used a 2x2 PD matrix, with positive reinforcements for T and R outcomes and negative reinforcements for P and S outcomes (Fig. 4). The positive payoffs were 1 or 4 chocolate rice chips weighting 200 mg each and the

negative payoffs were tail pinches delivered by grabbing the rat's tail and pinch it 1 or 3 times with a forceps in the distal third of the tail. The number of chocolate rice chips and tail pinches each subject received during each trial depended on the combination of both player choices in the matrix.

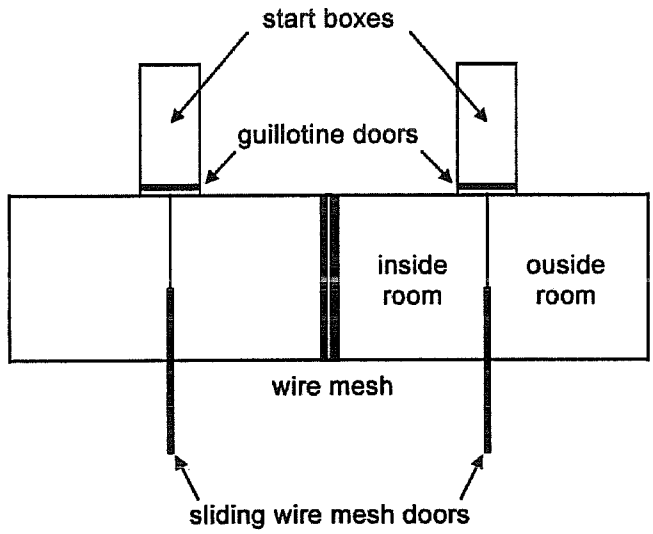


**Figure 3.** Diagram of the apparatus used in the experiments.

| | | Player 2 | |
|---|---|---|---|
| | | Cooperate | Defect |
| Player 1 | Cooperate | R= +1 | S= -3 |
| | Defect | T= +4 | P= -1 |

**Figure 4.** Game matrix used in the experiments. The values are referent to the row player. The positive values correspond to the number of chocolate rice chips and the negative values correspond to the number of given tail pinches.

### 2.4. General procedure

Experiments were conducted with naïve rats as subjects playing a similar general procedure. An experiment consisted of a variable number of daily sessions, conducted 7 days per week and each session containing 20 trials.

A trial was divided into two phases:

i) Decision phase - In the beginning of each trial the subject was placed inside the start box. The guillotine door of the start box was then opened and the subject entered in the arms of

the T-maze. The subject had 30 s to make a choice for a room. We considered a decision when the subject's body was completely inside one of the rooms. The sliding mesh grid was closed immediately after the subject's decision. After the 30 s period the subject was forced to make a decision by closing the sliding mesh grid slowly against the subject.

ii) Payoff phase - Payoffs were delivered directly in the chosen room and approximately 3-5 s after the closure of the door by opening the ceiling door. After payoff deliverance the subjects stayed undisturbed in the room for 30 s. The subject was then removed and a new trial started with a new rat. The inter-trial time was not fixed and was approximately 5-7 min.

There were no restrictions on social interactions. The only cues available when behaving were from the locations and behaviour of the subjects themselves without the aid of non-social feedback such as a light cues. This loosens the tight experimental control over individual behaviour that characterizes physical isolation models in favour of a context within which subjects develop their own idiosyncratic ways of satisfying the outcome contingency.

## 3. Results

### 3.1. Experiment 1 – Testing side bias

To test a putative side bias of the rats towards a specific room we used a simple procedure in which 4 subjects could explore freely the two rooms of the T-maze during 5 min. along 3 sessions in the absence of a conspecific and of any reinforcement delivery. In a following experiment we tested other 8 subjects during 5 min. along 9 sessions, again without any reinforcement delivery, but in the presence of a conspecific in the contiguous T-maze. Both the subject and the demonstrator could roam freely between the arms of their T-maze.

In the free run trials where rats could move freely between the two rooms without the presence of a conspecific, the subjects stayed in average 54,3 ± 4,63 % of the time in the inside room, not significantly different of an equivalent permanence in the outside room ($t=0,92$; $p=0,38$).

The percentage of time in the inside rooms across sessions when a conspecific was present in the other T-maze is shown in Figure 5. The subjects have shown a preference for staying in the inside room (immediately contiguous to the conspecific's T-maze) during the first seven sessions. However this preference was dissipated in the last two sessions. In the last session, the subjects stayed in average 52,0 ± 3,26 % of the time in the inside box, a percentage not significantly different of an even permanence in each box ($t=0,82$; $p=0,48$).

These results suggest that rats do not present a significant bias towards a given side of the T-maze and that an influence of the presence of a conspecific could be dissipated through familiarization when no reinforcements are delivered.

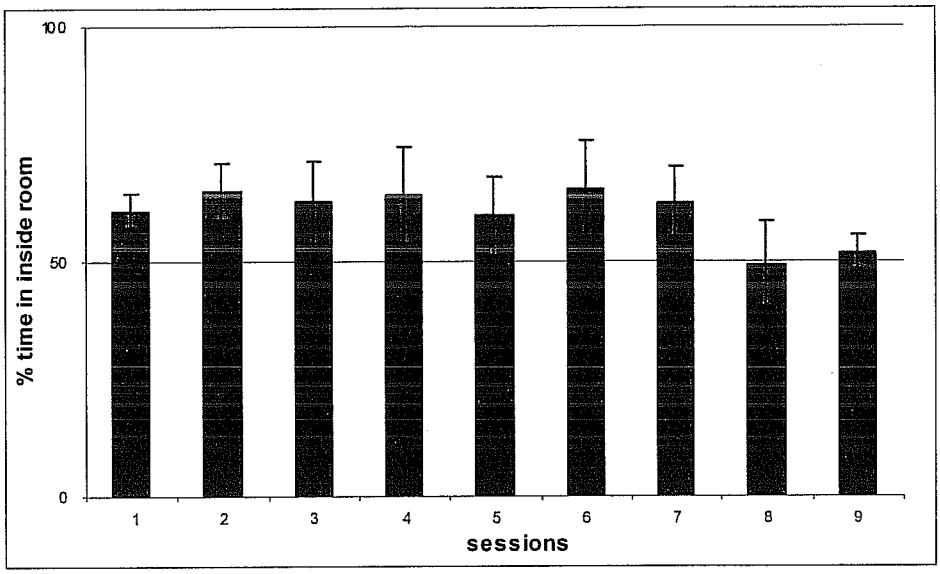


**Figure 5.** Mean percentage of time that rats stayed in the inside room of the T-maze when a conspecific was present in the contiguous T-maze. Dispersion bars represent Standard Errors.

### 3.2. Experiment 2 – Testing positive and negative reinforcements

To test if the could discriminate between the two positive reinforcements and the two negative reinforcements that form the game matrix a pair of experiences with a variable number of sessions were conducted. In a first one, four rats were tested using 1 or 4 chocolate chips delivered after a decision. Two of the subjects received the maximum reward in the inside room and the other two in the outside room. The daily sessions endured until rats acquired more than 80% of maximum reward decisions in two consecutive sessions. In the following sessions, the reinforcement values assigned to each room were reverted. An analogous experiment was conducted with other 4 rats using 1 or 3 tail pinches as negative reinforcements. When the minimum punishment decisions accounted for more than 80% of a daily session during two sessions the values assigned to the rooms were reverted.

Figures 6 and 7 show the average proportion of "correct" decisions of the rats during the experiments to test the discrimination between the two positive reinforcements and the two negative reinforcements as given by the game matrix. We assumed "correct" decisions as entering in the rooms that bear the most positive reward and the least negative punishment. In both tests, subjects quickly learn to maximize payoffs, taking respectively, 4 and 5 sessions to rise above 80% of correct decisions. However, the subjects respond differently to the reverse condition: The rats submitted to the reward reverse condition swiftly learn the new contingencies, taking 3 sessions to reach 80% of "correct" choices. On the other hand, the rat's choices were sub-optimal in the punishment reverse condition, ranging 69% of "correct" choices after 8 sessions.

Rats quickly learn to discriminate between two quantitatively distinct reinforcements showing a clear preference for entering in the rooms associated to high rewarding reinforcements and low punishing reinforcements, validating the T>R>P>S inequality proposed by our game matrix.

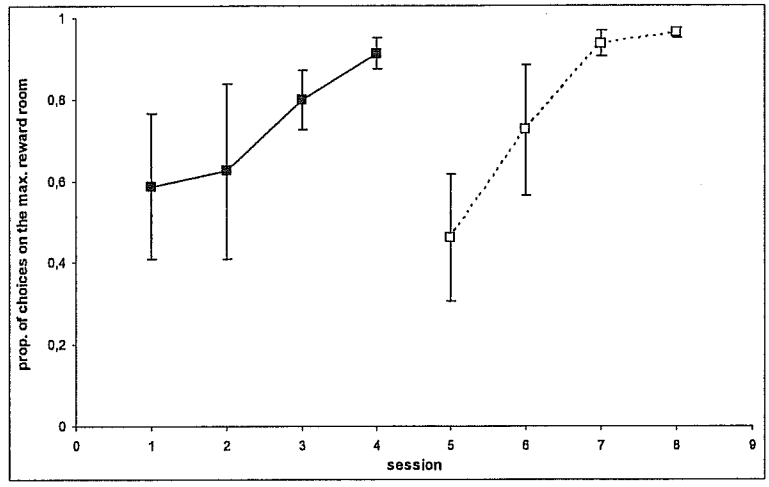


**Figure 6.** Average proportion of maximum reward room's choices. The rewards were 4 vs. 1 chocolate chips. The black line represent a first condition where for two subjects the maximum reward room was assigned to the inside room and for other two were assigned to the outside room. The dashed line represents the reverse condition. Dispersion bars represent Standard Errors.

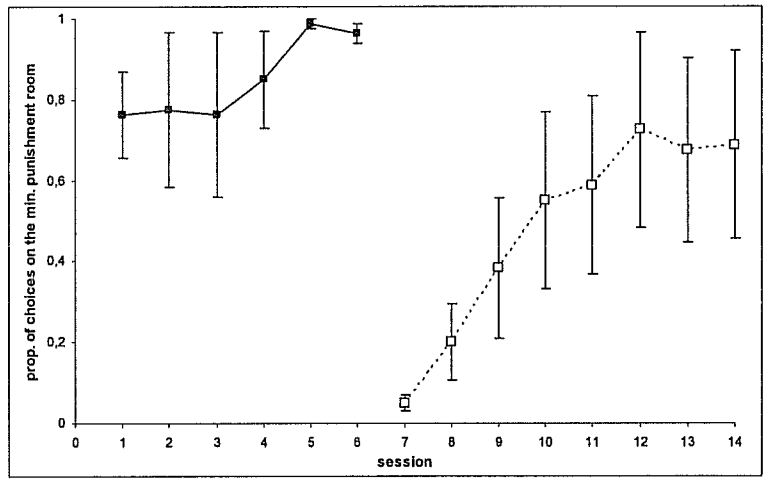


**Figure 7.** Average proportion of minimum punishment room's choices. The punishments were 3 vs. 1 tail pinches. The black line represent a first condition where for two subjects the minimum punishment room was assigned to the inside room and for other two were assigned to the outside room. The dashed line represents the reverse condition. Dispersion bars represent Standard Errors.

### 3.3. Experiment 3 – Rats vs. Pseudorandom and TFT strategies

Using the proposed game matrix, two strategies were run against rats in two separate experiments: Pseudorandom (PR) and Tit-for-Tat (TFT). In these, rats repeatedly choose between cooperation (C) and defection (D). To manipulate strategies, we assigned one and always the same rat to act as a demonstrator. The demonstrator was simply allocated in the room assigned for C or D in the beginning of each trial and the subject ended it. Four rats were used against each strategy, in which 2 were assigned to cooperate in the inside room and the other 2 the converse. The demonstrator was always assigned to C in the inside room and D in

the outside room. Each game lasted nineteen sessions. To minimize social effects, the subjects were submitted to a preliminary familiarization with the apparatus and the demonstrator during 9 sessions of 5 min where the subjects could roam freely between rooms.

In the PR condition, the demonstrator was randomly assigned to cooperate or defect on each trial, but regarding that i) within a session the demonstrator was assigned 10 Cooperation decisions and 10 Defection decisions and; ii) the highest number of consecutive C or D assignments were three. Playing against a Random strategy in an IPD game is equivalent to play a sum of single Prisoner's Dilemma games (Hall, 2003). Since the equilibrium in a single PD game is mutual defection, we expect that rats playing against a Random strategy show high levels of mutual defection along sessions.

In the TFT condition, the demonstrator choice on a trial was simply the subject own previous choice on the preceding trial, with the exception that on the first trial of a session the demonstrator always cooperated. According to Axelrod (1984) TFT properties: "niceness", "swift retaliation", and "quick forgiving", we expect that rats playing against this strategy present a certain but undetermined level of sustained cooperation through reciprocity.

In rats playing against the PR opponent, the mean proportion of cooperative responses decreased across sessions. Kendall's $\tau$ statistic, used to test for temporal trends (Fisch, 2001), indicated that this trend was statistically significant ($K=-117$; $p<0,005$). After the $11^{th}$ session, the proportion of cooperation reached 0,25 and maintained a consistent decrease until the end of the experiment, reaching a minimum value of 0,04 in the $18^{th}$ session.

Rats playing against a TFT opponent did not showed a consistent and statistically significant change in the cooperative choices across sessions ($K=21$; $p>0,05$), however the average proportion of cooperation rise constantly above 50% following the second session and reached the maximum value of 0,79 in the $5^{th}$ session. In half of the remaining sessions this proportion overcome 0,6 (Fig. 8).

A cumulative plot of the average proportion of cooperation was designed in order to access if, and in which session, the subjects reached an equilibrium response against the PR and TFT strategies (Fig. 9). This method provided a better estimate of the rat's preference in the experiments, while allowing a posterior sequential trial-by-trial analysis. The cumulative proportion of cooperation of subjects that played against a PR strategy did not showed stability even after 19 sessions, always decreasing across sessions, which suggest that the theoretical optimal response (always defect) towards this strategy was not fully reached. The cumulative plot of subjects that played against TFT increased swiftly in the first six sessions, followed by slight fluctuations until reached a stable equilibrium after the $11^{th}$ session at 0,59.

These results showed that rats playing against a TFT strategy early adopted a stable proportion of cooperative responses above the 50% and reached a steady equilibrium after 11 sessions, while rats playing against the PR showed a constant decrease in cooperative responses along sessions towards the theoretical equilibrium of 100% defection.

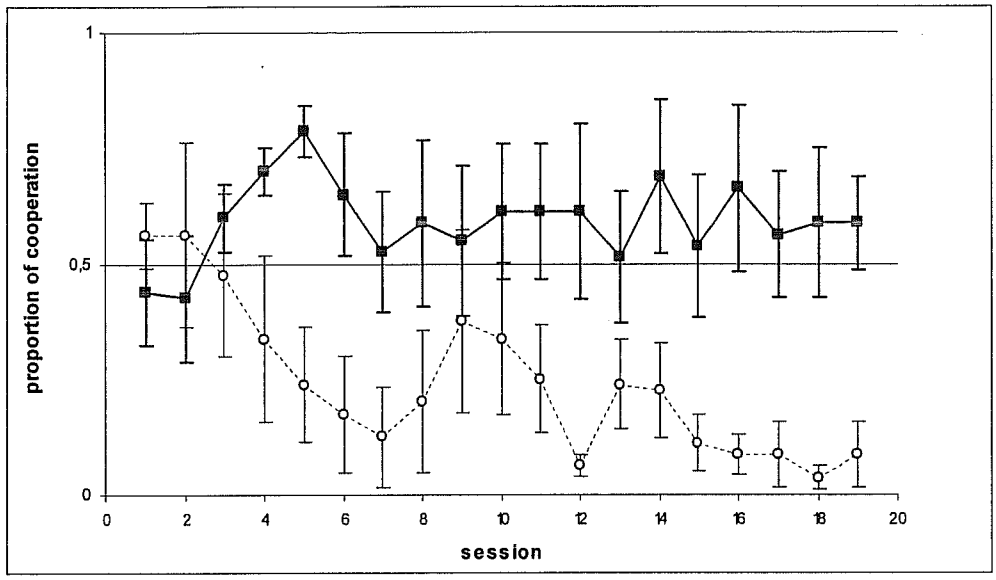


**Figure 8.** Average proportion of cooperation choices of rats against a Pseudorandom (dashed line) and a Tit-For-Tat (black line) strategy. Dispersion bars represent Standard Errors.

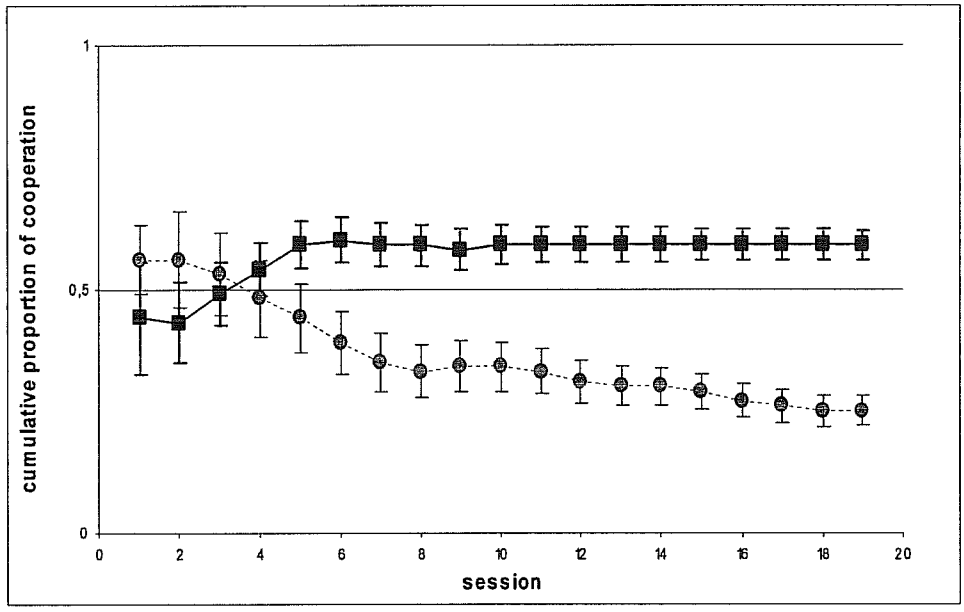


**Figure 9.** Cumulative average of the proportion of cooperation choices of rats against Pseudorandom (grey line) and Tit-For-Tat strategy (black line) strategies. Dispersion bars represent Standard Errors.

Following the cumulative average analysis we pooled the average proportion of cooperation of the last eight sessions of the PR and TFT conditions (Fig. 10). These two treatments were significantly different ($t=152,17$; $p<0,001$), with subjects playing against TFT ($0,59 \pm 0,04$) having a superior proportion of cooperation than those that played against PR ($0,12 \pm 0,04$). A one-sample T-test conducted showed that the proportion of cooperation of rats playing directly against TFT was significantly different of 0.5, value that represents a subjects decision made purely by chance ($t= 4,465$; $p<0,05$).

The global proportion of cooperation revealed that rats are able of a sustained cooperation above random levels if the opponent reciprocates (TFT opponent) when a game matrix with two qualitatively distinct reinforcements is used in a IPD context. Such cooperation levels are clearly higher from those obtained by rats playing against a Pseudorandom strategy, in which the theoretical equilibrium is always defect.

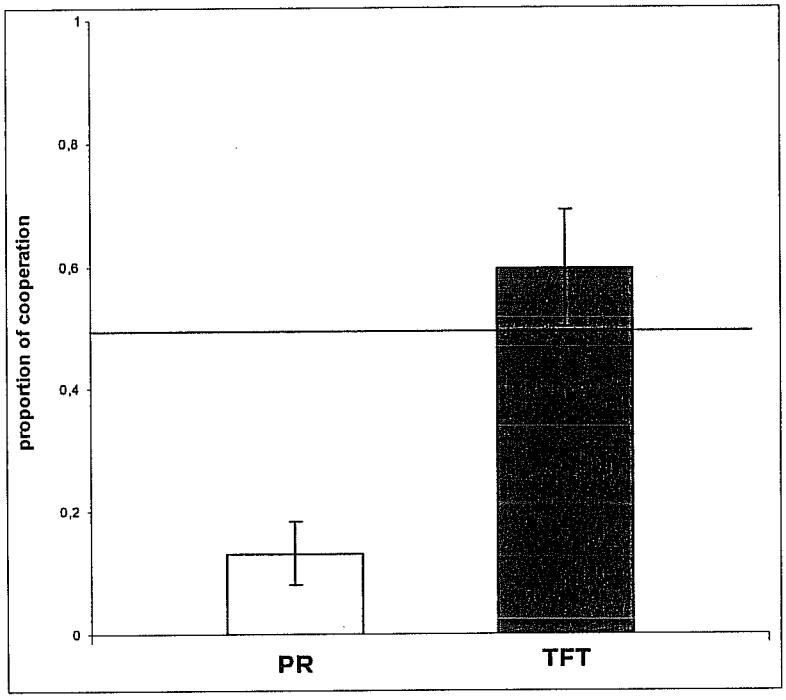


**Figure 10.** Mean proportion of cooperation in the last eight sessions of the PR and TFT conditions. Bars represent 95% confidence limits.

Figure 11 shows the mean number of positive and negative reinforcements received by rats across sessions. In rats that played against a PR opponent the maximization of the payoffs due to the increase on defection across sessions was associated to a clear raise in the number of positive reinforcements received (chiefly via an increase of T outcomes) ($K=91$; $p<0,05$) and a more gradual, but statistically significant, decrease in the number of negative reinforcements received ($K= -89$; $p<0,05$) (mainly by a decrease of S outcomes). On the other hand, reinforcements received by the rats playing against TFT were stable across sessions ($K= -21$; $p>0,05$ and $K= -27$; $p>0,05$, for positive and negative reinforcements, respectively), with the positive reinforcements delivered mainly by R outcomes and the negative reinforcements by S outcomes.

These results showed that rats playing against a PR opponent tend to adopt an optimal strategy along sessions, with a rapid increase in the number of positive reinforcements and a

gradual decrease in the number of positive reinforcements received. Rats that played against a TFT opponent, albeit showing sustained levels of cooperation, seems to adopt a sub-optimal strategy across sessions, with less positive reinforcements received and similar levels of negative reinforcements than the subjects that played against PR.

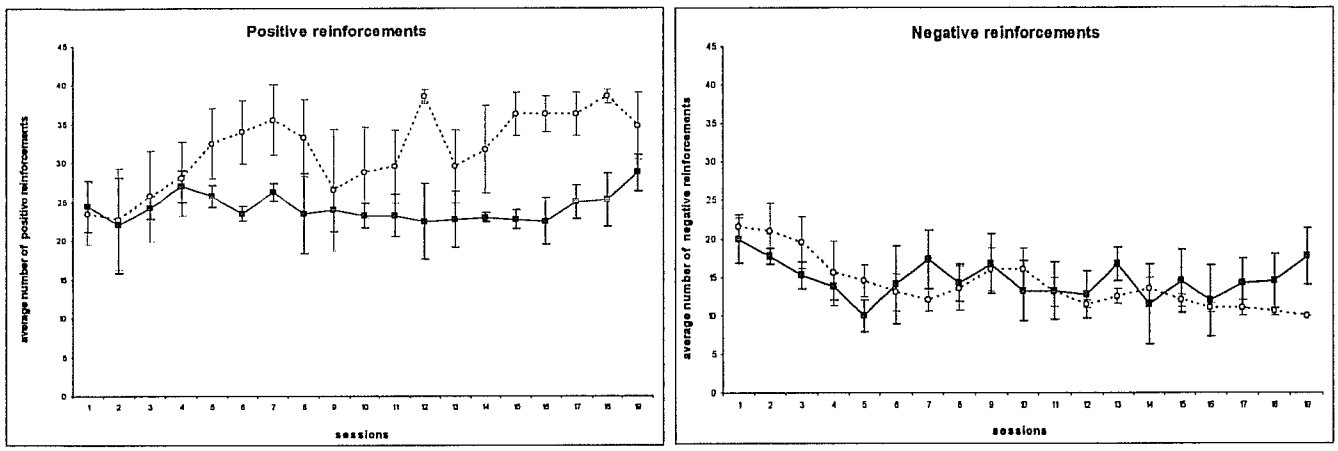


**Figure 11.** Average number of positive and negative reinforcements received by rats across sessions against the Pseudorandom (dash line) and Tit-For-Tat (black line) strategies. Dispersion bars represent Standard Errors.

The local strategic (trial-by-trial) decisions of rats playing against a TFT opponent were analysed using the data of the last eight sessions. Table 1 shows the probability of cooperating following a previous outcome (Stephens et al., 1997, 2002). For example, *t* represents the probability of cooperating in the trial after an outcome of T. The overall observed strategy was *r*>*t*≈*p*>s; that is, rats against TFT were more likely to cooperate after mutual defection and least likely to cooperate after being "suckered", and cooperated at roughly equivalent intermediate levels after defecting.

The bottom rows of the table represent the theoretical strategy probabilities for a pair of Tit-For-Tat and Pavlov strategists. If two players played TFT, we would expect high *t* and *r* values and low *p* and *s* values. That is, cooperation should follow previous opponent cooperation (T and R trials). Pavlov predicts the "win-stay, lose-shift" strategy, in which players will repeat a rewarding choice and switch after a punishing choice. Therefore, we would expect high *r* and *p* values and low *t* and *s*. The rats, however, diverged markedly from both theoretical strategies. For example, both theoretical strategies predict that *s*=0 (do not cooperate with a player that just "suckered" you), but the subjects were forgiving.

**Table 1.** The mean probability of the subject cooperating in the trial following the T, R, P, and S payoffs was calculated for the TFT experiment. The values for theoretical TFT and Pavlov represent predicted strategy probabilities for subjects implementing those strategies.

| | *t* | *r* | *p* | *s* |
|---|---|---|---|---|
| TFT opponent | 0,58 | 0,76 | 0,58 | 0,50 |
| Theoretical TFT | 1 | 1 | 0 | 0 |
| Theoretical Pavlov | 0 | 1 | 1 | 0 |

Lag 1 state transition diagrams of the same data were analysed and Cramer's $\Phi$ correlation coefficients calculated for all possible transitions. This measure of association converts the $\chi^2$ (in which the expected values were considered to be random) of all state transitions into a coefficient that varies from 0 to 1 (Lehner, 1996). The computed $\Phi$ coefficients are indexed to the likelihood that the subjects will switch or stay in each of the four outcomes, R, S, T and P. Note that only certain transitions are possible when a subject plays against TFT, given that the demonstrator always copies the subject's previous move. Note also that when playing against TFT, the subject begins a session either at R or at T, given that the demonstrator always cooperates on the first trial. Figure 12 shows that the higher $\Phi$ values were for R→R and P→P transitions, which indicate that subjects were more likely to sustain mutual cooperation and mutual defection in the proceeding trial. After a S outcome, rats were more likely to switch to defection and receive T than they were to continue to cooperate and receive R.

These trial-by-trial analyses show that rats that played against TFT employed a sub-optimal strategy not related with known theoretical strategies. However, the observed strategy is consistent, showing high indicators of sustained cooperation and defection and high levels of "forgiveness".

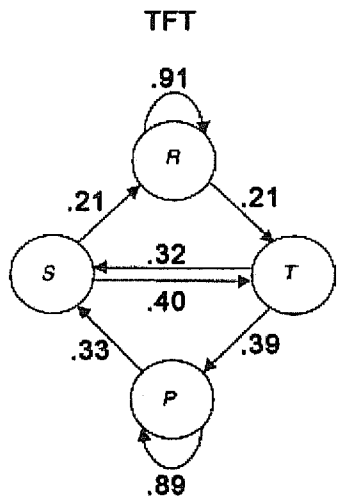


**Figure 12.** Lag 1 state transition diagrams showing the likelihood that the rats received a particular outcome (R, S, T or P) on trial given that they received one of these outcomes on the preceding trial. The numbers next to each arrow are Cramer's $\Phi$ correlation coefficients. Transitions are shown for the last eight sessions.

To assess if the observed proportion of cooperation were biased due to individual options of rats across sessions in the PR and TFT conditions, the individual performance of subjects that were assigned to cooperate away or near the demonstrator were plotted (Fig. 13). A visual inspection reveals that the PR condition rats assigned to cooperate in the inside room, and hence, nearer of the conspecific, defected less when compared to the rats that were assigned to cooperate in the outside room. In the TFT condition this trend was less clear, but the less cooperative rat (markedly less than the other 3 subjects) was assigned to cooperate away of the conspecific.

The observed differences in the individual performances of rats in the experiment appear to be related with the influence of the presence and position in the T-maze of a conspecific in the subject's decisions.

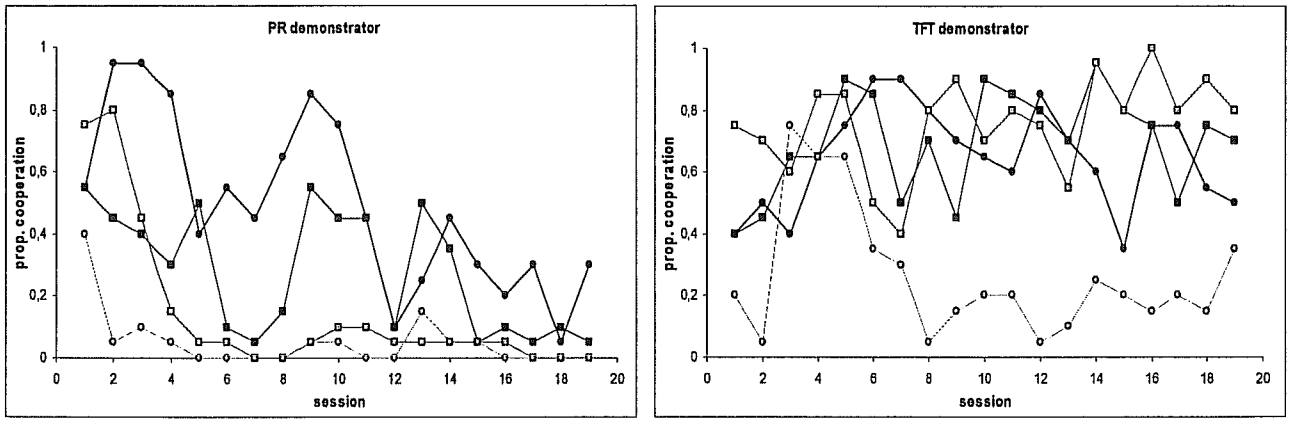


**Figure 13.** Individual performance of the rats against a PR and a TFT demonstrator. The black lines represent subjects assigned to cooperate in the inside room and the converse for subjects symbolized by grey lines.

### 3.4. Experiment 4 – The influence of the demonstrator in subject's decisions

To test if the differences observed between rats assigned to cooperate near or away the conspecific may be due to the influence of the presence of a conspecific in the subject's decisions towards positive and negative reinforcements a pair of experiences were conducted.

In a first experiment, we investigated the subject's decisions between a large positive reinforcement (4 chocolate chips) away of a conspecific versus a small positive reinforcement (1 chocolate chip) in the proximity of the conspecific. In the second experiment, we investigated the subject's decisions between a small negative reinforcement (1 tail pinch) away of a conspecific versus a bigger negative reinforcement (3 tail pinches) in the proximity of the conspecific. Both experiments had two treatments: a *near* condition, in which the conspecific was placed in the adjacent proximity of the subject's T-maze and an *away* condition, in which the conspecific was placed in the farthest room of the subject's T-maze. In each of these 4 treatments 4 rats were tested during 5 sessions (Fig. 14).

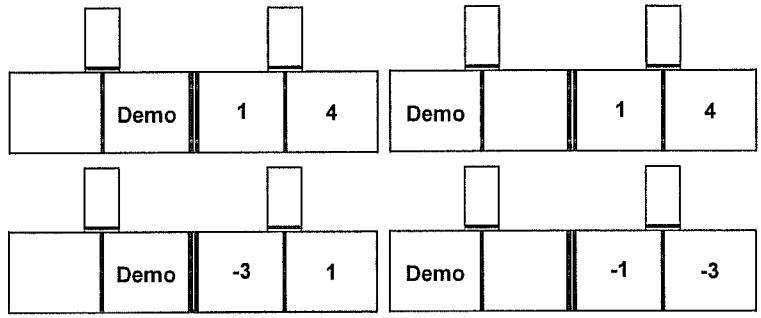


**Figure 14.** Experimental design to test the influence of the presence of a conspecific demonstrator in the rats decisions towards positive and negative reinforcements.

The rats assigned to chose between 1 chocolate chips in the adjacent presence of a conspecific vs. 4 chocolate chips far of its presence initially chose to the maximum reward room in similar proportion on both *away* and *near* treatments (0,34 ± 0,08 and 0,40 ± 0,007, respectively). In the fifth and last session, however, the subjects that experienced the *away* condition chose preferably the room farthest of the conspecific with the larger positive reinforcement (0,90 ± 0,01), while the subjects on the *near* treatment showed a significantly preference towards the room immediately adjacent to the conspecific (0,38 ± 0,11) ($t=4,4$; $p=0,004$) (Fig. 15). The similar experience using negative reinforcements, in which the rats had to choose between 3 tail pinches in the adjacent presence of a conspecific vs. 1 tail pinch far of its presence showed similar responses of the subjects between treatments, either in the initial session (0,53 ± 0,11 and 0,50 ± 0,12, for the *away* and *near* treatments, respectively; $t=1,48$; $p>0,05$) and in the final session (0,93 ± 0,01 for both *away* and *near* treatments; $t\approx0$;$p\approx1$) (Fig. 16). The *no demonstrator* treatment data was obtained from the rats that received highest reward and the highest punishment in the inside room on the payoff discrimination experiences, hence $n=2$ for this treatment. The final proportion of subjects under this condition was 0,98 ± 0,01 for the reward experiment and 0,98 ± 0,02 for the punishment experiment.

These results show that the conspecific presence and position in the apparatus influenced the rats' decisions towards a given room. Rats showed a marked preference to stay in the adjacent room of the conspecific, even receiving a lower positive reinforcement. This preference was null when the conspecific was away of the subject and when the delivered reinforcements were negative.

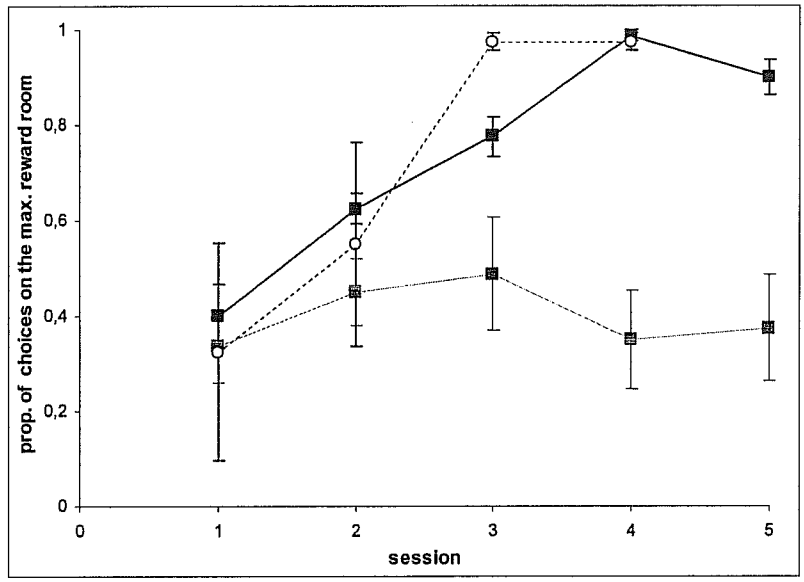


**Figure 15.** Average proportion of maximum reward decisions in the presence of a demonstrator during 5 sessions. The black line represents the *away* condition, in which the demonstrator was allocated in the farthest room of the subject's T-maze. The grey line represents the *near* condition, in which the demonstrator was allocated in the room immediately adjacent to the subject's T-maze. For both these conditions, n=4. The dashed line represents a *no demonstrator* condition, n=2.

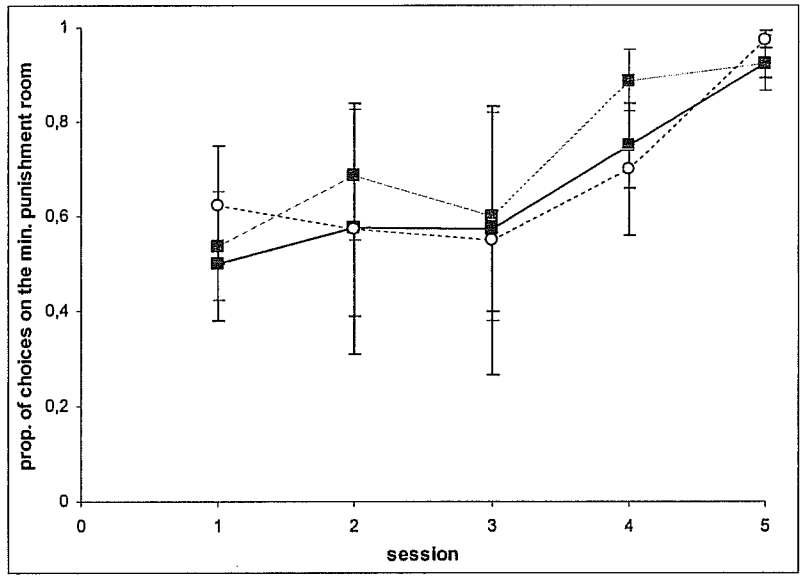


**Figure 16.** Average proportion of minimum punishment decisions in the presence of a demonstrator during 5 sessions. The black line represents the *away* condition, in which the demonstrator was allocated in the farthest room of the subject's T-maze. The grey line represents the *near* condition, in which the demonstrator was allocated in the room immediately adjacent to the subject's T-maze. For both these conditions, n=4. The dashed line represents a *no demonstrator* condition, n=2.

### 3.5. Experiment 5 – The role of prior contingencies in rats vs. TFT strategy

To test the role of prior contingencies in the subject's response against a TFT opponent, 4 rats played against a PR demonstrator during 19 sessions following by other 19 sessions against the same demonstrator but this time using a TFT strategy. The experimental details are similar to the ones described in the preceding experiment. This TFT condition is termed in future references as (PR)TFT.

The rats that played against a Pseudorandom strategy following by a swift to a Tit-For-Tat strategy, showed in the PR phase a decrease in the proportion of cooperative choices ($K=-109$; $p<0,005$) similar to the anterior PR experiment. The proportion of cooperation reached 0,25 in the 9$^{th}$ session and maintained the decrease until the end of the first phase, reaching a minimum value of 0,13 in the 18$^{th}$ session (Fig. 17).

In the 20$^{th}$ session the conspecific demonstrator switch to a TFT strategy: this led to a small increment in the proportion of cooperation in the first (PR)TFT session (from 0,16 in the last PR session to 0,25 in the first TFT session). Nevertheless, in the following sessions the proportion of cooperation remained stable ($K=-31$; $p>0,05$) around 0,26 (Fig. 17). The pooled variance of the first 5 sessions of the (PR)TFT condition was significantly higher than the variance of the last 5 sessions of the PR condition ($F=0,32$;$p<0,01$). Both results suggest that the rats experienced the contingency switch and altered their decisions, at least in the first (PR)TFT sessions.

The pooled average of the proportion of cooperation in the last eight (PR)TFT sessions was 0,26 ± 0,09, significantly lower than the cooperation levels observed in rats that played directly against TFT ($t=12,6$; $p<0,001$). These results showed that rats although experiencing the strategy switch from PR to TFT, responded sub-optimal to the second condition when compared with subjects that played directly against TFT, which demonstrate that rats' decisions under a IPD game are sensible to past reinforcement contingencies.

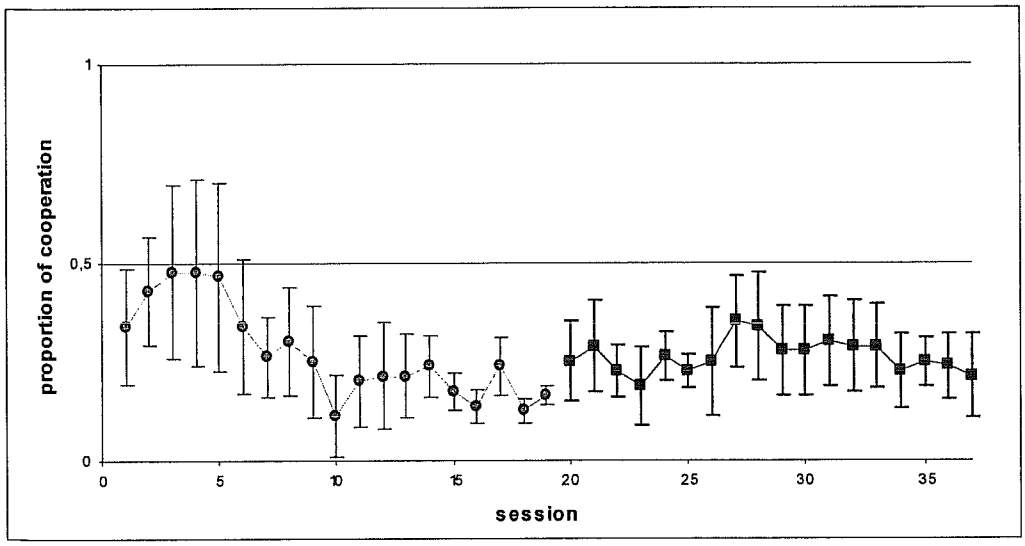


**Figure 17.** Mean proportion of cooperation choices of rats against a Tit For Tat strategy (black line) prior a Pseudorandom strategy (grey line).

## 4. Discussion

The results of the present study demonstrate that when their choices are reciprocated (against a Tit-For-Tat opponent), rats can sustain cooperation during an Iterated Prisoner's Dilemma game (Fig. 8). This agrees with the theoretical conclusions of Stephens and Clements (1998) model, in which a IPD game matrix with S and P outcomes i) unambiguously and universally punishing and ii) qualitatively distinct from the T and R consequences, lead to high levels of cooperation. A similar level of cooperation (about 60%) was observed in humans that used game matrices with money earnings in the T and R outcomes and money losses in P and S outcomes (Rapoport and Chammah, 1965).

The game matrix used in this present study is comparable to that used in human studies in the sense that rats deal with chocolate chips and tail pinches, as humans treat losses and gains as qualitatively distinct entities (Kahneman and Tversky, 1979). The utilization of this kind of matrix appears to relax the instability of rats' cooperative behaviour observed in previous studies (Flood et al. 1983; Gardner et al., 1984), not by diminishing their preferences for immediate benefits, which Stevens, Cushman and Hauser (2005) suggested as being a major factor influencing cooperative equilibria, but rather because some degree of cooperation is required to avoid the physical consequences of negative payoffs (Stephens and Clements, 1998).

Quite surprisingly, the utilization of a qualitatively different game matrix was never intentionally accounted in non-human animal IPD studies, though most of the described dilemmas that animals face in wild have qualitatively different immediate consequences. This difficulty was turned around in some naturalistic studies by converting different consequences, like starvation and satiation, into a common currency (usually fitness or probability of survival) (e.g. Wilkinson, 1984). However, such procedure could be seen as an artefact, transforming a non-IPD situation with immediate consequences in an IPD context in which the consequences happened in the long run and vice-versa (Nöe, 2006).

According to the classic IPD theoretical structure (Axelrod and Hamilton, 1981), some reservations exist in consider a game with such game matrix a truly Prisoner's Dilemma game because the condition $R>(T+S)/2$ is arithmetical nonsense (Stephens and Clements, 1998). However, in a strict economical sense, and as in humans (Rapoport and Chammah, 1965), rats are able to categorize different payoffs according the first PD condition, that is, they rank the outcomes in the sense that T payoff (4 chocolate chips) is preferred to R (1 chocolate chip) is preferred to P (1 tail pinch) is preferred to S (3 tail pinches). A detach of the classical IPD structure, by allowing the categorization of qualitatively different consequences (Stephens and Clements, 1998), could represent a plausible model of how animals face the cooperative dilemmas in nature. Future lab studies should focus in how animals weight and categorize such qualitatively different consequences towards an improved IPD paradigm.

This work resumes the laboratory research on cooperation using rats as models within a Prisoners' Dilemma game initiated in the 80's. Our general conclusions challenge the results obtained by Flood et al. (1983) and Gardner et al. (1984) that found low levels of sustain cooperation (between 30% and 40%, respectively). The cooperation levels found by Flood et al. (1983) could be due the employment of a negative game matrix (different time delays until same food deliverance). The use of differential time delays in the construction of a PD matrix was indirectly contested by Stevens et al. (2005) which suggested that temporal discounting is a cognitive constrain in the emergence of cooperation by reciprocity and not a reinforcement perceived by the animals. Another criticism to Flood work was that animals that decided but the adversary didn't, the first received the shorter of the two payoffs associated to its decision. That procedure signifies that such animals sensed the inaction of the adversary as a non cooperative act. Such procedure represents a real modification of the game matrix that could explain the low cooperative levels in these rats.

On the other hand, Gardner et al. (1984) using a classical positive matrix (food pellets) found levels of cooperation around 40% with a Plexiglas partition separating the T-mazes. Although visual information about the partner's choices was apparently important in the subjects' decisions in an IPD, two factors could concurred in the observed low levels of cooperation: these authors possibly over emphasized the value of visual information on the rats' decisions, and didn't allow that players used other sensory modalities in obtaining information about the game partner. And nost importantly, the classical matrix used possibly did not elicit players to cooperate since the differential consequences were not perceived by the subjects.

It's possible that the innovative game matrix used in this present work could elicit higher levels of cooperation in rats. However, this assumption needs to be test by submitting pairs of rats to a truly IPD game. Some authors suggest that two-player IPD's are (almost) as rare in the laboratory as in nature by stating the argument that a truly IPD situation only exist when controlling the flow of information between the players, while proving that the subjects are still aware of each other's presence and of each other's influence on payoffs, and know which choice the partner has made (Nöe, 2006). This claim seems to be unfounded and an overuse of the metaphor that is associated with the Prisoner's Dilemma game: In fact, the only requirements necessary to define a situation as a PD game are the economic relation between the payoffs. Nevertheless, the management of information, the perception of the conspecific as a player and a sensible quantification of the payoffs as perceived by subjects are crucial when investigate new suitable paradigms that could explain cooperation trough reciprocity. In this sense, future experiments should also take in account the value of sensorial information within a game context, for instances, by using an opaque partition separating pairs of rats playing IPD, and thus replicating Gardner et al. (1984) methodology. Several alternative experiments blocking the role of olfactive and acoustic information between players should also be considered. Further studies should also maintain and improve the sequential analyses of the trial-by-trial data, which could enlighten how and why rats implement and maintained game strategies against a theoretical strategy or against a truly IPD opponent.

Deciding in entering in one of two rooms in a laboratory apparatus seems an unlikely act of cooperation. One would not suppose that natural selection has favoured cooperative room entering in T-mazes. However, it is reasonable to suppose that natural selection has equipped animals with the ability to recognize and implement cooperative strategies, in novel situations, when economic circumstances favour cooperation.

Trial-by-trial analyses revealed that rats that played directly against TFT showed a stable game strategy that albeit consistent, disagrees with both Pavlov and TFT theoretical strategies. These subjects seems to adopt a game strategy with high likelihood to sustain mutual cooperation or defection, while showing low levels of "forgiveness", that is, after being "suckered" rats reciprocate immediately with defection. High positive reinforcements, surprisingly, elicited a similar likelihood of subsequent cooperation and defection. Albeit, such strategy and rats seem with high levels of sustain cooperation (high *r,* as predicted for a TFT or a Pavlov player), but "forgiving" (s>0, different of both theoretical strategies), and revealing roughly equivalent cooperation in the two "mixed" situations (T and S outcomes). An equal game strategy was previously observed in jays playing against a TFT conspecific opponent using a positive reinforcement game matrix (Stephens et al., 2002).

The presence and position of a conspecific opponent inside the experimental apparatus appears to influence the rats' decisions towards a given room (Figs. 15 and 16). Since being with a conspecific can be a reward in itself, such "external payoffs" can cause considerable discrepancy between payoffs intended by the experimenters and payoffs as perceived by the subjects (Nöe, 2006). Although being hard to quantify such social influence, this study weighs up the external payoffs that are associated with a presence of a conspecific in the reinforcement matrix used. Even though this analysis was not made in an IPD game context, the data obtained revealed some interesting conclusions:

i) The conspecific presence appears to have a null value when the rats had to decide between two different negative reinforcements;

ii) However, there is an observable effect of the conspecific presence when rats had to decide between two positive reinforcements.

Our results corroborates some laboratory experiments using coordination tasks in rats (Schuster and Perelberg, 2004) and non-human primates (Mendres and de Waal, 2000; Hauser et al., 2003) which showed that when given a choice to accomplish a task alone or in the company of a conspecific, subjects preferred the former even if this was associated to lower food reinforcements. The external payoffs associated to the presence of a conspecific appear to be sensible to local contingencies, and in this sense, modify the cooperative tendencies and game strategy of rats playing an IPD game. In a wider perspective, a simple economical analysis of the animals' decisions under a game context is insufficient to converge empirical observation with the classical IPD theoretical framework.

The reinforcement value associated to the conspecific presence also poses challenges for further investigation. The main problem resides in the fact that these external payoffs appear

to have different values depending on the quality of the reinforcement. Two possibilities could be explored to improve the proposed game matrix: A first approach relies on the quantification of the external payoffs values for each of the four possible outcomes within an IPD game and the posterior rearrange of the game matrix by adding or subtracting these values to the payoffs, in order to assure that the game matrix still obeys PD specifications. A second and mutual exclusive approach relies on the experimental extinction of the social bias. Since our results suggest that the influence of external payoffs only is observable when positive reinforcements are delivered, an alternative could be replace the chocolate chips to a highly rewarding reinforcement like drops of sucrose solution. Such reward is well known to be very rewarding to rats and simultaneously could avoid the problems faced due the neophobia that rats showed towards the chocolate chips.

Our results also reveal that rats are sensible to past reinforcement contingencies in the sense that subjects submitted to a previous Random strategy showed significantly lower levels of cooperation when playing against TFT strategy than rats that played directly against TFT. These results were striking different from Hall (2003) study in which pigeons playing against a TFT opponent did not show differences in cooperation levels and slopes across sessions either beginning playing against a Random opponent or directly to a TFT opponent. The divergence between these two results could be explained by different methodological and game matrices approaches: First, Hall pigeons' showed colour bias towards the buttons that represented "cooperation" and "defection"; second, a transitional procedure between conditions was implemented that was not entirely effective in "washing out" the colour bias; and third, there were difficulties in follow the proposed matrix (T=5, R=3, P=1, S=0) due to inaccuracies in measuring the exact quantity of food consumed by animals. Clements and Stephens (1995) study showed that blue jays were sensible to past reinforcement contingencies when switching game matrices: pairs of birds were exposed to an iterated game that started with an Prisoner's Dilemma matrix, then switch to a mutualistic matrix, and than returned to the first matrix. The matrix change was determined by the percentage of mutual defection or cooperation across sessions. Although jays in IPD contexts rapidly reached mutual defection, they took significantly more sessions to reached defection after experiencing the mutualistic matrix (in which the equilibrium for maximize gains is mutual cooperation).

In conclusion, the present study demonstrates that non-human animals could cooperate by reciprocity under an Iterated Prisoner's Dilemma game by applying a game matrix composed with positive and negative reinforcements. However, sustained cooperation and game strategies are apparently affected by social effects and past reinforcement contingencies. Such implies that an exclusively economical perspective could be insufficient to explain the cooperative behaviours of animals. Iterated Prisoner's Dilemma games could still present a valid tool for the empirical study of animal cooperation if one takes in account the possible social and historical contingencies intrinsic to cooperative interactions between animals.

## 5. Acknowledgments

This master thesis was supported by an "Instituto Gulbenkian da Ciência" grant and was made under the project "Prisioner's Dilemma in Rats". The candidate wishes to thank to Marta Moita for the deposited trust and for the scientific supervision and orientation, to Élio Sucena and Isabel Gordo for additional scientific supervision and valuable inputs and comments and a special word for Iris Vilares, not only a co-worker in the project, but mainly a generous friend that with constant support and sharp mind made this work possible and better.

## 6. Bibliography

Axelrod R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Axelrod R., & Hamilton W. (1981). The evolution of cooperation. *Science.* **211**: 1390-1396.

Baker, F., & Rachlin, H. (2002). Self-control by pigeons in the prisoner's dilemma. *Psychonomic Bulletin & Review.* **9 (3)**: 482-488.

Bowles, S., & Gintis, H. (2003). The Origins of Human Cooperation. *The Genetic and Cultural Origins of Cooperation.* ed Hammerstein, P. MIT Press:Cambridge.

Brown J. (1983). Cooperation - a biologist's dilemma. *Advances in the Study of Behaviour.* ed. Rosenblatt J. S. , pp. 1–37. New York Academic.

Clements, K., & Stephens, D. (1995). Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. *Animal Behaviour.* **50**: 527–535.

Connor R. (1995a). The benefits of mutualism: a conceptual framework. *Biological Reviews.*1-31.

Connor, R. (1995b). Altruism among non-relatives: alternatives to the 'Prisoner's Dilemma'. *Trends in Ecology and Evolution.* **10**: 84–86.

Connor R. (1996). Partner preferences in byproduct mutualisms and the case of predator inspection in fish. *Anim. Behav.* **51**: 451– 454.

Darwin C. (1859). *On the origin of species.* ed. Murray J. London.

Davies N., & Houston A. (1981). Owners and satellites: the economics of territory defence in the pied wagtail *Motacilla alba. Journal of Animal Ecology.* **50**: 157-180.

de Waal F., & Berger M. (2000). Payment for labour in monkeys. *Nature.* **404**: 563.

Doebeli, M., & Hauert, C. (2005). Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecology Letters.* **8**: 748–766.

Dugatkin L. (1988). Do guppies play Tit for Tat during predator inspection visits? *Behav. Ecol. Sociobiol.* **23**: 395–399.

Dugatkin L. (1991a). Predator inspection, TIT FOR TAT and shoaling: a comment on masters and waite. *Animal Behaviour.* **41**: 898-900.

Dugatkin L. (1991b). Dynamics of the TIT FOR TAT strategy during predator inspection in guppies. *Behavioral Ecology and Sociobiology.* **29**: 127-132.

Dugatkin L. (1997). The Evolution of Cooperation. *Bioscience.* **47 (6)**: 6-20.

Dugatkin L. (2002). Animal cooperation among unrelated individuals. *Naturwissenschaften.* **89**: 533–541.

Dugatkin L., & Alfieri M. (1991). Guppies and the TIT FOR TAT strategy: preference based on past interaction. *Behavioral Ecology and Sociobiology.* **28**: 243-246.

Dugatkin L., & Alfieri M. (2001). A cognitive approach to the study of animal cooperation. *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, ed. Bekoff M., Allen C., Burghardt G. M., pp. 413–419. Cambridge, MA: MIT Press.

Dugatkin L., Mesterton-Gibbons M., & Houston A. (1992). Beyond the prisoner's dilemma: towards models to discriminate among mechanisms of cooperation in nature. *Tr. Ecol. Evol.* **7**: 202–205.

Dugatkin L., & Reeve H. (1998). *Game theory and animal behavior.* New York: Oxford University Press.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *TRENDS in Cognitive Sciences.* **8 (4)**: 185-190.

Fisch, G. (2001). Evaluating data from behavioral analysis: visual inspection or statistical models. *Behavioural Processes.* **54**: 137-154.

Flood, M., Lendenmann, K., & Rappoport, A. (1983). 2x2 games played by rats: different delays of reinforcement as payoffs. *Behav. Sci.* **28**: 65–78.

Galef Jr., B., & Whiskin, E. (2003). Preference for Novel Flavors in Adult Norway Rats (*Rattus norvegicus*). *Journal of Comparative Psychology*. **117 (1)**: 96-100.

Gardner, R., Corbin, T., Beltramo, J., & Nickell, G. (1984). The prisoner's dilemma game and cooperation in the rat. *Psychol. Rep.* **55**: 687–696.

Gardner, A., & West, S. (2004) Cooperation and Punishment, Especially in Humans. *The american naturalist*. **164 (6)**: 753-764.

Gintis, H. (2000). Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology* **206**:169–179.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior* **24**: 153–172.

Green, L., Price, P., & Hamburger, M. (1995) Prisoner's Dilemma and the pigeon: control by immediate consequences. *Journal of the Experimental Analysis of Behavior*. **64**: 1–17.

Grutter, A. (1999). Cleaner fish really do clean. *Nature*. **398**: 672-673.

Hall, S. (2003). Transitions between cooperative and non cooperative responding in the 'Pigeon's Dilemma'. *Behavioural Processes*. **60**: 199-208.

Hamilton W. (1964). The genetical evolution of social behaviour. I, II. *J. Theor. Biol.* **7**: 1–52.

Hart B., & Hart L. (1992). Reciprocal allogrooming in impala, *Aepyceros melampus. Anim. Behav*. **44**: 1073–1083.

Hauber M., & Sherman P. (2001). Self-referent phenotype matching: theoretical considerations and empirical evidence. *Trends Neurosci.* **24**: 609–616.

Hauser M., Chen M., Chen F., & Chuang E. (2003). Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proc. R. Soc. London. Ser. B.* **270**: 2363–2370.

Harris A., & Madden G. (2002). Delay discounting and performance on the prisoner's dilemma game. *Psychol. Rec.* **52**: 429–440.

Hammerstein, P., & Hagen, E. (2005). The second wave of evolutionary economics in biology. *TRENDS in Ecology and Evolution.* **20 (11)**: 604-609.

Henrich, J., & Henrich, N. (2006) Culture, evolution and the puzzle of human cooperation. *Cognitive Systems Research.* **7**: 220–245.

Kahneman, D. and Tversky, A. (1979) Prospect Theory: An analysis of decision under risk. *Econometrica.* **47**: 263-291.

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000) Social exchange and reciprocity: confusion or a heuristic. *Evol. Hum. Behav.* **21**: 411–427.

Kropotkin (1902) *Mutual Aid; a factor of evolution.* Project Gutenberg.org

Lazarus J., & Metcalfe N. (1990) Tit-for-tat cooperation in sticklebacks: a critique of Milinski. *Anim. Behav.* **39**: 987–988.

Lehner, P. (1996) *Handbook of ethological methods, second ed.* Cambridge University Press, Cambridge.

Levine, D. (1998) Modeling altruism and spitefulness in experiments. *Rev. Econ. Dynam.* **1**: 593–622.

Masters W., & Waite T. (1990). Tit-for-tat during predator inspection or shoaling? *Anim. Behav.* **39**: 603–604.

Mazur J. (1987). An adjusting procedure for studying delayed reinforcement. *Quantitative Analyses of Behavior: the Effect of Delay and of Intervening Events on Reinforcement Value.* ed. Rachlin H., pp. 55–73. Hillsdale, NJ: Erlbaum.

Mendres, K., & de Waal, F. (2000). Capuchins do cooperate: the advantage of an intuitive task. *Animal Behaviour.* **60**: 523–529.

Milinski M. (1987). TIT FOR TAT in sticklebacks and the evolution of cooperation. *Nature.* **325**: 433–435.

Milinski M. (1996). Is by-product mutualism better than tit-for-tat reciprocity in explaining cooperative predator inspection. *Animal Behaviour.* **51**: 458-461.

Milinski M., & Wedekind C. (1998). Working memory constrains human cooperation in the Prisoner's Dilemma. *Proc. Natl. Acad. Sci. USA.* **95**: 13755–13758.

Nöe R. (2006). Cooperation experiments: coordination through communication versus acting apart together. *Animal Behavior.* **71**: 1-18.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms Tit-for-Tat in the Prisoner's Dilemma game. *Nature, Lond.* **364**: 56-58.

Osborne, M., & Rubinstein, A. (1994). *A course in game theory.* The MIT Press: Cambridge.

Posch, M. (1999). Win-Stay, Lose-Shift Strategies for Repeated Games: Memory Length, Aspiration Levels and Noise. *J. theor. Biol.* **198**: 183-195.

Rachlin H., Raineri A., & Cross D. (1991). Subjective probability and delay. *J. Exp. Anal. Behav.* **55**: 233–244.

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Rapoport, A. and Chammah, A. (1965) *Prisoner's Dilemma: A study in Conflict annd Cooperation.* Ann Arbor, MI: University of Michigan Press.

Richards J., Mitchell S., deWit H., & Seiden L. (1997). Determination of discount functions in rats with an adjusting-amount procedure. *J. Exp. Anal. Behav.* **67**:353–366.

Reboreda, J., & Kacelnik, A. (1993) The role of autoshaping in cooperative two-player games between starlings. *J. exp. Analysis Behav.* **60**: 67–83.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002) A neural basis for social cooperation. *Neuron* **35**: 395–405.

Rissing S., & Pollock G. (1991) An experimental analysis of pleometric advantage in *Messor pergandei. Ins Soc.* **63**:205–211.

Roberts, G. (1997). Testing mutualism: a commentary on Clements & Stephens. *Anim. Behav.* **53**: 1361–1362.

Sachs, J., Mueller, U., Wilcox, T., & Bull, J. (2004). The evolution of cooperation. *Quarterly Review of Biology.* **79**: 135–160.

Schuster, R., & Perelberg, A. (2004). Why cooperate? An economic perspective is not enough. *Behavioural Processes.* **66**: 261–277.

Semmann, D., Krambeck, H., & Milinski, M. (2004). Strategic investment in reputation. *Behav Ecol Sociobiol* **56**: 248–252.

Sherman P. (1977). Nepotism and the evolution of alarm calls. *Science* **197**: 1246-1253.

Sherman P. (1985). Alarm calls of Belding's ground squirrels to aerial predators: nepotism or self-preservation? *Behavioral Ecology and Sociobiology* **17**: 313-323.

Stephens, D., & Anderson, J. (1997). Reply to Roberts: cooperation is an outcome, not a mechanism. *Anim. Behav.* **53**: 1363-1364.

Stephens D., Anderson J., & Benson K. (1997). On the spurious occurrence of Tit for Tat in pairs of predator approaching fish. *Anim. Behav.* **53**: 113–131.

Stephens, D., & Clements, K. (1998). Game Theory and Learning. *Game theory and animal behavior.* (Eds.) Dugatkin L., Reeve H.. New York: Oxford University Press. pp: 239-260.

Stephens D., McLinn C., & Stevens J. (2002). Discounting and reciprocity in an Iterated Prisoner's Dilemma. *Science* **298**: 2216–2218.

Stephens D., McLinn C., & Stevens J. (2006). Effects of temporal clumping and payoff accumulation on impulsiveness and cooperation. *Behavioural Processes.* **71**: 29-40.

Stevens, J., & Stephens, D. (2004) The economic basis of cooperation: tradeoffs between selfishness and generosity. *Behavioral Ecology.* **15**: 255–261.

Stevens J., Cushman F., & Hauser M. (2005) Evolving the Psychological Mechanisms for Cooperation. *Annu. Rev. Ecol. Evol. Syst.* **36**: 499–518.

Thornidike, E. (1911). *Animal Intelligence.* New York.

Trivers, R. (1971). The evolution of reciprocal altruism. Q. Rev. Biol. 46: 35–57.

West Eberhard, M. (1975). The evolution of social behavior by kin selection. Q. Rev. Biol. 50: 1–33.

Wilkinson, G. (1984). Reciprocal food sharing in the vampire bat. Nature 308: 181–84.

Wilson, E. (1975). Sociobiology: The New Synthesis. Cambridge, MA: Harvard Univ. Press.

Wilson, D., & Sober, E. (1994). Re-introducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences* .**17**: 585-654.